(2)

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

ARTIFICIAL NEURAL NETWORKS AND THEIR APPLICATIONS IN DIAGNOSTICS OF INCIPIENT FAULTS IN ROTATING MACHINERY

by

David K. Carlson

March 1991

Thesis Advisor:	Y.S. Shin
Thesis Co-Advisor	D.S. Kim

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b RESTRICTIVE MARKINGS

2a SECURITY CLASSIFICATION AUTHORITY

2b DECLASSIFICATION / DOWNGRADING SCHEDULE

3 DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release;
distribution is unlimited

4 PERFORMING ORGANIZATION REPORT NUMBER(S)

5 MONITORING ORGANIZATION REPORT NUMBER(S)

6a NAME OF PERFORMING ORGANIZATION
Naval Postgraduate School

6b OFFICE SYMBOL (If applicable)
Code 34

7a NAME OF MONITORING ORGANIZATION
Naval Postgraduate School

6c ADDRESS (City, State, and ZIP Code)
Monterey, CA 93943-5000

7b ADDRESS (City, State, and ZIP Code)
Monterey, CA 93943-5000

8a NAME OF FUNDING / SPONSORING ORGANIZATION

8b OFFICE SYMBOL (If applicable)

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c ADDRESS (City, State, and ZIP Code)

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
ARTIFICIAL NEURAL NETWORKS AND THEIR APPLICATIONS IN DIAGNOSTICS OF INCIPIENT FAULTS IN ROTATING MACHINERY

12 PERSONAL AUTHOR(S)
Carlson, David K.

13a TYPE OF REPORT
Master's Thesis

13b TIME COVERED
FROM ________ TO ______

14 DATE OF REPORT (Year, Month, Day)
March 1991

15 PAGE COUNT
198

16 SUPPLEMENTARY NOTATION The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defence or the U.S. Government.

17 COSATI CODES

FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Machinery condition monitoring, Machinery vibration monitoring, Machinery diagnostics, Neural networks, Backpropagation

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

In an effort to curtail rising operating costs, machinery condition monitoring and diagnostics are being increasingly used as a part of predictive maintenance programs. Vibration analysis is currently among the most effective tools in machinery condition monitoring and diagnostics but has proven difficult to automate fully. Artificial Neural Networks, patterned after neurological systems, provide a heuristic, data-based approach to problems and have demonstrated robust behavior when faced with unique and noisy data. Thus neural networks may provide an alternative or complement to conventional rule based expert systems in machinery diagnostics applications. Research is presented wherein a series of neural networks utilizing the highly successful backpropagation paradigm are configured to provide machinery diagnostics for comparitively uncomplicated mechanical systems. Through observation of their responses to minor architectural changes and performance upon presentation of genuine and

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a NAME OF RESPONSIBLE INDIVIDUAL
Y.S. Shin

22b TELEPHONE (Include Area Code)
(408) 646-2568

22c OFFICE SYMBOL
Code 69Sg

DD Form 1473, JUN 86 Previous editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

S/N 0102-LF-014-6603

Block 19. artificially generated vibration data, an effort is made to ascertain their utility in more complicated systems.

Approved for public release; distribution is unlimited.

Artificicial Neural Networks and Their
Applications in Diagnostics of Incipient
Faults in Rotating Machinery

by

David K. Carlson
Lieutenant , United States Navy
B.S., U.S. Merchant Marine Academy, Kings Point, New York

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN MECHANICAL ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
March 1991

Author: ______________________
David K. Carlson

Approved by: ______________________
Y.S. Shin, Thesis Advisor

____
D.S. Kim, Thesis Co-advisor

____
for A.J. Healey, Chairman
Department of Mechanical Engineering

ABSTRACT

In an effort to curtail rising operating costs, machinery condition monitoring and diagnostics are being increasingly used as part of predictive maintenance programs. Vibration analysis is currently among the most effective tools in machinery condition monitoring and diagnostics but has proven difficult to automate fully. Artificial Neural Networks, patterned after neurological systems, provide a heuristic, data based approach to problems and have demonstrated robust behavior when faced with unique and noisy data. Thus neural networks may provide an alternative or complement to conventional rule based expert systems in machinery diagnostics applications. Research is presented wherein a series of neural networks utilizing the highly successful backpropagation paradigm are configured to provide machinery diagnostics for comparatively uncomplicated mechanical systems Through observation of their responses to minor architectural changes and performance upon presentation of genuine and artificially generated vibration data, an effort is made to ascertain their utility in more complicated systems.

TABLE OF CONTENTS

ACKNOWLEDGEMENT

I wish to express my sincere appreciation to Professor Young Sik Shin for his guidance, patience, and support throughout the course of this thesis research. I would also like to thank Professor Dong Soo Kim for his long hours and tireless help and support particularly in the technical aspects of completing this thesis. Additionally, I would like to thank the Mr. Alan Pride, Project Manager, Submarine Noise and Vibration Monitoring of the Navsea Submarine Monitoring, Maintenance, and Support Program Office, without whose support and funding this project would not have been possible. Finally I would like to express my appreciation for the support and encouragement of Mrs. Anne Stevens and her daughter Carrie, who provided a home away from home during the limited hours diverted away from the thesis and academics throughout my tenure at the Naval Postgraduate School.

I. INTRODUCTION

As operating costs continue to rise, greater emphasis on minimizing down time of critical machinery by establishing effective machinery maintenance programs. By far the most efficient of the major maintenance programs available is the corrective maintenance program. The critical factor in implementing this program is a reliable means by which to monitor the health of operating machinery and to diagnose the source of the fault when something goes wrong. While this has traditionally been accomplished by highly capable and qualified machinery experts, their small number and expense makes it highly desirable to automate the machinery monitoring and diagnostics process. Indeed there have been a number of rule based expert systems placed on the market in an effort to satisfy this need. Unfortunately they have not proven entirely successful. Principal areas of weakness lie in the nature of the problem. Mathematical characterization of all but the most elementary mechanical systems exceeds current computational capability. The sources of mechanical excitation include multiple sources of noise which tend to confuse conventional rule based expert systems. Often the nature of mechanical vibration troubleshooting does not conduce itself well with the series nature of conventional computers.

Artificial Neural Networks possess features that may help alleviate a number of these characteristic problems. Neural networks are data-based vice rule based, thereby possessing the potential of being able to operate where analytical solutions are inadequate. They are reputed to be robust and highly tolerant of noisy data. They are parallel in nature which gives them certain advantages in assimilating the experience of existing biological "expert systems" in ways completely different from the manner in which current expert systems must operate.

While Artificial Neural Networks have only come into their own since 1985, they are not entirely untried. Neural Networks have been assimilated into a number of engineering applications. In the Chemical Engineering field, Watanabe and Himelblau[Ref.1] as well as Venkatasubramanian and Chan[Ref.2] have utilized multi-layered neural networks to assist in chemical process fault diagnostics. In the Medical Engineering field, Porenta et al[Ref.3] developed a pattern recognition system which identified diseased and healthy coronary arteries based on scintigram profiles and Iwata et al [Ref.4] developed a data compression system to increase the recording capacity of Holter portable EKG machines. In the Automotive Industry Marko et al [Ref.5] developed a neural network based diagnostic system for use with an electronic engine control computer. In the Aeronautical Engineering field, McDuff, et al [Ref.6] developed an engine fault detection system utilizing

an ART1 learning algoıithm, while Dietz, Kiech and Ali [Ref.7] developed a similar device for the F/A 18 using the backpropagation learning algorithm. This is only a few of the applications currently in progress. Application in machinery condition monitoring and diagnostics is a logical extension.

This paper is broken up into six additional sections. The remainder of this section further elaborates on the background, intentions, and direction of this research. Chapter II provides a brief overview of the theory and development of artificial neural networks and particularly the backpropagation paradigm. Chapter III provides background information on machinery diagnostics. Chapter IV describes a series of preliminary experiments on which a prototype neural network diagnostics models was based and includes a sensitivity analysis of the neural networks to the number of processing elements in its hidden layer. Chapter V presents the physical model for which the prototype neural networks diagnostics models were designed and describes the empirical data acquisition process. Chapter VI describes the architecture, training methodology , and responses to empirical and artificially generated data for the prototype neural network diagnostics models.

A. MACHINERY MAINTENANCE PROGRAMS

All industrial organizations utilizing any range of mechanical equipment will tend to schedule the maintenance of

that equipment in accordance with one or several of the three following general machinery maintenance programs. The simplest and least efficient of these programs is a corrective maintenance program. Here the equipment is allowed to operate without any intervention by service personnel until it breaks down, whereupon the equipment is serviced to correct the casualty and then returned to operation. This maintenance program has the advantages of being easy to manage and inexpensive to implement until the equipment breaks down. Its drawbacks are that once the equipment does break down, the damage suffered by the equipment is likely to be severe and the attendant down time extensive. Furthermore, the equipment breakdown will be unscheduled and will have an adverse effect on the operation of the entire plant should the equipment not be redundant and still be essential to the plant's operation. This has the tendency to make this machinery maintenance program prohibitively expensive in all but the least sophisticated operations.

Preventive maintenance consists of a managed program of periodic maintenance checks scheduled throughout the service life of the machinery. The periodicity of these checks is generally based on corporate experience with the more sophisticated checks and those requiring extensive down time occurring much less frequently than less sophisticated checks and those requiring little or no down time. This program requires considerably more management and involves

considerably more intervention by service personnel than the corrective maintenance program and is correspondingly more expensive to implement. However, although the frequency of short down periods for the equipment increase, the long down times and great expense associated with catastrophic failures is substantially reduced. Further, the down times for the equipment can be efficiently scheduled to minimize interference with plant operation whereas the down periods associated with the corrective maintenance program could not. This aspect of a preventive maintenance program is its chief attraction and preventive maintenance programs have achieved widespread acceptance throughout industry and government.

Preventive maintenance is not without its drawbacks, however. Often the corporate experience associated with a particular machinery component is limited and, to compensate for this, periodicities for the various checks are compressed. While this may not be a problem with maintenance checks requiring minimal down time, financial outlay, or technical expertise, there are numerous checks that do require significant outlays of these scarce resources and thus contribute to the inefficiency of plant operation. Further, even with the best preventive maintenance program, equipment will break down unexpectedly on occasion, albeit at a much reduced rate than that found in a corrective maintenance program. Preventive maintenance can also give rise to self-imposed casualties. Scarcely an experienced technician exists

who has not encountered a situation where a previously smoothly operating machine has undergone a maintenance check following which it has broken down due to some error in reassembly. While ensuring that the experience level of those conducting the maintenance check is appropriate to its complexity will reduce the number of these occurrences, it will never completely alleviate them.

A predictive maintenance program, where the health of machinery components could be determined while in an on-line status and component faults could be predicted well in advance of failure would allow for timely and scheduled correction of faults without requiring unnecessary and expensive maintenance checks. This type of program would be ideal, providing all of the benefits of both corrective and preventive maintenance programs without their attendant drawbacks. However, this program would have to include a highly reliable means of machinery fault prediction in order to be successful. To accomplish this a reliable means of machinery condition monitoring and diagnostics must be obtained.

B. MACHINERY CONDITION MONITORING AND DIAGNOSTICS

To be successful a machinery condition monitoring system must be capable of obtaining the required information about the machinery while it is in an on-line status. Currently numerous system-wide operating parameters are methodically monitored either manually or with automated data recording

systems, but in general, the data obtained by these means while sufficient to monitor the system or plant as a whole are insufficient to determine the status of components of individual machines to the point of providing the basis for an effective predictive maintenance program. Three fields of condition monitoring that show promise in providing such detailed information include temperature analysis, tribology, and vibration analysis. However, while detailed temperature analysis is limited to machinery involved in a thermal cycle, and tribology requires a means of extracting machinery wear products from the machine such as a lube oil filter, vibration analysis can be used on any machine involving moving parts without interrupting that machine's operation and has the potential to provide the detailed information required to reliably predict machinery faults well in advance of failure.

Since its inception, great progress has been made in the field of vibration analysis. Analytical solutions for the most elementary mechanical systems have been in existence for a long time. As improvements in computer-based modal analysis techniques continue to be made, the level of complexity of mechanical systems that can be solved by numerical and analytical means improves correspondingly. Nevertheless, the extreme complexity of existing and anticipated mechanical systems, as well as the physical limitations of sensor placement, the presence of extraneous noise, and transient operation complicate the machinery vibration problem to the

point that it is doubtful that analytical or numerical methods will be able to provide practical solutions to real machinery diagnostics problems.

This does not invalidate the utility of vibration analysis in the field of machinery condition monitoring and diagnostics. Experienced technicians have long astounded engineers by their ability to predict and identify machinery faults merely by listening to and touching their machinery. By combining heuristic and analytical knowledge with modern vibration monitoring instrumentation, a significant machinery diagnostic capability has been achieved. However, to be reliable, this analysis has had to be conducted by a limited number of experts. The rapid rise of computer technology has somewhat alleviated the problem of too few machinery diagnostics experts through the proliferation of rule based expert systems. However complicated series of IF-THEN statements are not always sufficient to accurately represent a knowledge base nor are they capable of easily incorporating new information as it becomes available. They are also generally less effective at detecting multiple faults than the experts that programmed them, and they are susceptible to error when provided partial or noisy information. Perhaps a data based approach rather than a rule based approach could help solve the limitations of conventional expert systems.

In the last several years a great deal of interest has been generated in a new branch of artificial intelligence

based on the theoretical operation of biological nervous systems. This branch of artificial intelligence features massively parallel networks of simple processing elements which function in a manner similar to biological neurons. These artificial neural networks learn the patterns associated with a given solution space by being provided a series of example vectors associated with that solution space. This data based vice rule based approach may make artificial neural networks a powerful tool in the field of vibration based machinery condition monitoring and diagnostics. A schematic of how the neural network would fit into the machinery condition monitoring and diagnostics scheme is provided in Figure 1.

C. INTENT AND DIRECTION OF RESEARCH

Artificial neural networks are gaining popularity in a number of applications including pattern recognition, signal processing, and non-linear optimization. The purpose and intent of this research is:

- To determine the feasibility of the application of artificial neural networks to machinery diagnostics by means of simple models and predominantly artificially generated data.
- To develop and a moderate complexity neural network model representing a physical model with multiple machinery components.
- To train and test this prototype neural network based machinery diagnostics model using both artificial and empirical data.

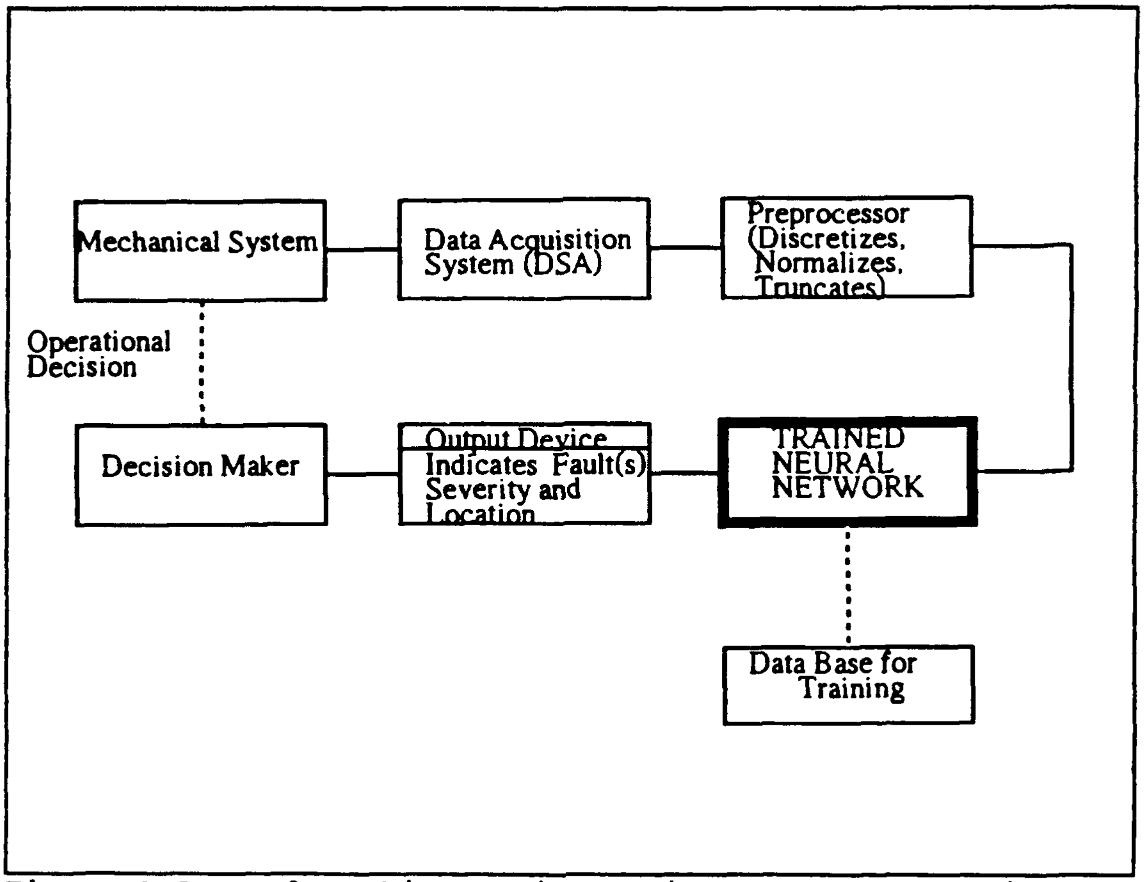

Figure 1 General Machinery Diagnostics System Schematic

- To ultimately incorporate neural networks into a diagnostic system for a highly complicated machinery system with highly transient operating conditions.

This thesis will focus primarily on the first three elements. However, the ultimate direction of focus of the research should also be kept in mind.

II. NEURAL NETWORK OVERVIEW

An Artificial Neural Network(ANN) is a massively parallel distributed processing system consisting of a series of interconnected individual processing elements which process information in a manner similar to that theoretically employed by neurons in biological systems.

In biological systems each neuron receives electrochemical stimulation from other neurons through its dendrites and axons by means of interneural connections called synapses. If the stimulation is sufficient, the individual neuron undergoes an electrochemical response and transmits this response to other neurons through various synapses. The strength of these synapses are as much a factor in determining the degree of excitation of the neuron as is the input stimulation itself.

Similarly, in ANN's, each processing element or artificial neuron is connected to several other processing elements by means of connections which are assigned a weighting of variable strength. The processing element then transmits a new signal to other processing elements depending on the value of a threshold as well as the strength of the input signal and the weighting of the connection. A schematic is provided in Figure 2.

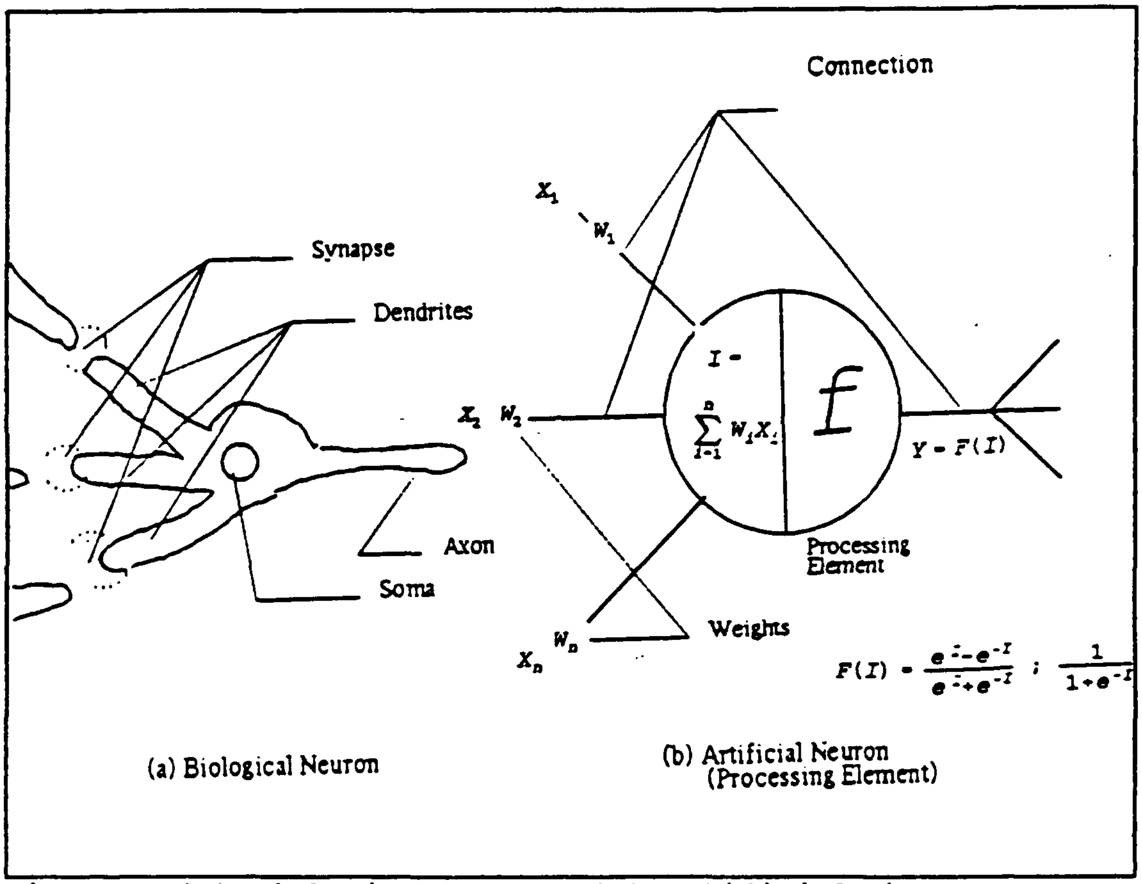

Figure 2 (A) Biological Neuron (B) Artificial Neuron

An ANN is generally composed of several levels of multiple processing elements, the lowest of which receives an input vector with one component of the vector introduced to each processing element. The responses of these processing elements are each transmitted to all processing elements of the next level, whose responses are in turn transmitted to each element of the following level. Thus the input vector is processed by each successive level of processing elements until the final level is reached. The response of this layer composes the output of the network. A schematic of this process is provided in Figure 3.

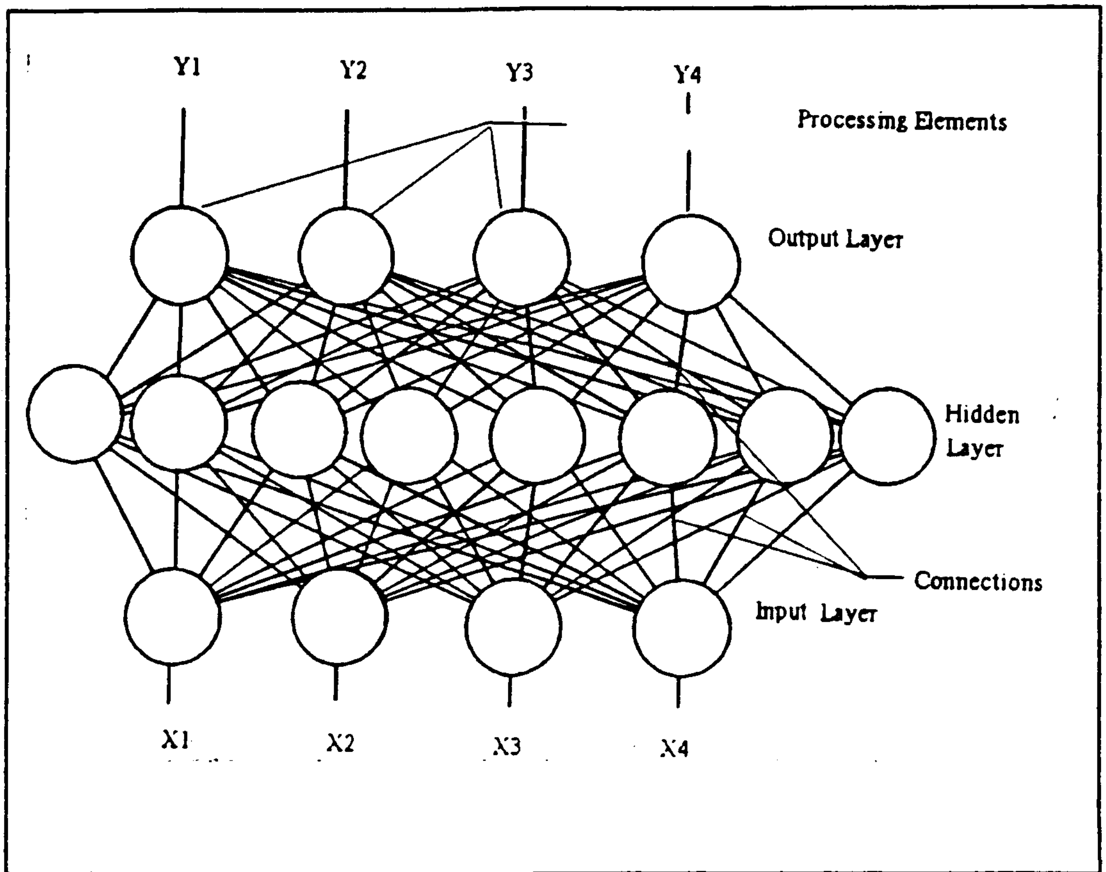

Figure 3 Generic Artificial Neural Network

This chapter is intended to provide the reader a brief overview of the terminology associated with neural networks, their history, and a synopsis of some of the learning algorithms and architectures currently being employed in neural computing. Particular attention will be given to the backpropagation algorithm as its use in machinery diagnostics is the focus of this research.

A. BASIC DEFINITIONS

1. Processing Element

A processing element(PE) is the lowest level self-contained computing element in the neural network. It typically is composed of three parts: a summer, a transfer function, and a threshold. The PE first sums all inputs it receives from outside the network or from other PE's. This sum is then compared to a threshold, which in several algorithms is zero. If the summed value is greater than the threshold, the summed value is processed by a generally non-linear transfer function. This non-linear transfer function is the heart of the processing element and gives the neural network the capability to discern non-linear relationships. It is also this transfer function that separates the artificial neural network from Bayesian nearest neighbors and statistical least squares approaches.

2. Layer

A layer is a group of PE's which are interconnected to other layers in the network but are not interconnected among PE's within their own layer. Layers are generally of three types: input, hidden, and output layers. PE's from the input layer are only connected to other PE's on the output side and receive input external to the network. PE's in the output layer are interconnected with other PE's on the input side and transmit output external to the network. Hidden

layers are intermediary layers consisting of groups of non-interconnected PE's which receive and transmit signals from other layers of PE's. The primary role of the hidden layer is to extract features from the previous layer for mapping to the next layer.

3. Connections

Connections are the means by which signals are transmitted throughout the network and are analogous to the dendrites and axons of the biological neuron. Each connection is a one or two-way path from one processing element to another. Each connection has a weight associated with it which is analogous to a synapse in a biological neural network. The values of the weights determine how the input vector maps onto the solution space and are the key instruments by which the network recognizes various patterns and relationships.

4. Learning

Originally the connection weights are established randomly throughout the network. The process by which the connection weights are adjusted to map the input vectors to the solution space is called "learning". There are two general types of learning. The first is supervised learning, where the weights are adjusted by some algorithm using a training set of input vectors. Here the actual output of the network is compared with a "target" or desired output and the connection weights are adjusted accordingly. The second type of learning

is unsupervised learning, where the network is left to itself to categorize various input vectors given an established threshold. This type of system is analogous to the statistical nearest neighbors classifier.

The key differences between various neural network architectures lies predominantly on the way they "learn". This is determined entirely by their learning algorithms, a few of which will be described shortly. However, a good deal of insight into the nature of neural networks can be obtained through a look at their developmental history.

B. HISTORY

The idea of creating a thinking machine based on biological learning theory gained momentum in the late 1940's when McCulloch and Pitts[Ref.9] published a paper "A Logical Calculus of Ideas Imminent in Nervous Activity", which stimulated interest in digital computers, a macroscopic rule-based approach to artificial intelligence, and biologically based artificial intelligence. Biologically based artificial intelligence gained further momentum when Hebb[Ref.10], a neurobiologist, formulated a means wherein neurons might learn, the Hebbian learning rule which was described earlier. This notion gained great public interest when in 1958 Rosenblatt[Ref.11] published research on an artificial neural network inspired by the optical pattern recognition capability of the eye based on processing elements called perceptrons.

Around 1960 Widrow and Hoff[Ref.12] developed an improved neural network based on the perceptron called an Adaline (Adaptive Linear Element) , which was the basis of the first commercially successful neural network enterprise, the Memistor corporation. They also developed a theorem which stated that an adaline and a perceptron are each capable of classifying any input space that could be linearly separated into two regions.[Refs.8 and 13]

The perceptron, however, for all its utility, had a critical drawback in that it required that the decision space be capable of being separated into two regions by means of a hyperplane. This drawback was criticized severely in Minsky and Papert's[Ref.14] book Perceptrons, where it was determined that the perceptron was incapable of solving the elementary exclusive OR logic problem. It was also criticized for not having a means to adjust weights in the case of incorrect outputs in multi-layer application. This criticism sharply reduced interest and funding in the biologically based artificial intelligence field.

Work continued in spite of little publicity and funding. In 1974 Werbos[Ref.15] completed a PhD dissertation that described an algorithm that provided a means to adjust perceptron weights in response to output errors that would eventually be improved upon and known as the backpropagation algorithm. Grossberg[Ref.16] continued work developing learning models based rigidly on neurobiological and learning

theory. In 1982 Hopfield[Ref.17] presented a paper on a neural computing model based on the olfactory system of garden slugs which built on previous work by Grossberg. This paper, presented by a widely respected scientist, renewed interest in neural computing. In 1986 Rumelhart improved upon the work of Werbos and developed the highly popular and successful backpropagation algorithm and, together with McClelland, Hinton and Williams[Ref.18], has continued to develop it. Since this time the field of neural computing has grown rapidly, with new applications being discovered regularly.[Ref.8]

The numbers and areas where applications for neural networks are being found span several disciplines and seem to focus on tasks such as signal processing, non-linear optimization, and pattern recognition. Their signal processing capability has been exploited in the medical field in the compression of electrocardiogram signals[Ref.4]; in image processing while subjected to noisy input data; and in predicting complicated series based on prior histories such as in weather prediction, general mathematics, and the stock market [Refs.8 and 19]. Their optimization capability has been exploited in determining optimum travel itineraries, circuit wiring, and non-linear control systems[Refs.8 and 19]. Their pattern recognition capabilities have been utilized in speech and symbol recognition[Refs.8 and 19], medical diagnostics[Ref.3], chemical processing[Refs.1 and 2], sonar

classification[Ref.20], electrical surge protection circuit testing[Ref.21], and engine fault detection[Refs.6 and 7]. This is a very limited listing of successful applications. Some of these have provided direct insights on how to approach the machinery diagnostics problem and will be described in later sections of this paper.

C. LEARNING RULES AND ARCHITECTURE

In conventional computing in general and in building expert systems in particular, the program software and rules formulated through collaboration of programming and subject experts is the heart of the system. In neural computing, the network architecture and learning algorithms used by the processing elements is central to the system. There are a great number of learning algorithms currently in use with some more popular than others for engineering applications.

1. Supervised Learning: General

Supervised learning can be subdivided into three general forms. These are Hebbian learning, Delta learning, and competetive learning.

a. *Hebbian Learning*

Hebbian learning is based on the premise that those connections that receive the most signal energy should in turn be strengthened.In this type of neural network, connection weights increase in a manner proportional to the magnitude of the signals provided that both the input through the path and

the desired output are high. While historically important and neurologically accurate, it is not widely used in neural computing applications.

b. Delta Rule Learning

Delta rule learning is probably the most popular type of learning currently in use. Here, weights are adjusted based on a direct comparison between the actual and desired outputs. Backpropagation is one learning rule based on the generalized delta rule:

$$W_{ij} = C_1 E_{ij} + C_2 M_{ij} + C_3 X_{ij} \tag{1}$$

where W_{ij} is the weight of the connection from the ith element in the current layer to the jth element of the previous layer; C_1, C_2, and C_3 are coefficients varying from 0 to 1; E_{ij} is the error proportional to the difference between the actual and desired output of the network; M_{ij} is the momentum term based on the difference between the previous weight of the given connection and the weight immediately prior to that; and X_{ij} is the activation energy associated with that particular connection.[Ref.8]

c. Competetive Learning

Competetive learning is where the output of processing elements is weighted according to the magnitude of its response relative to those of other processing elements. The "winning" processing element weighting is then modified according to a comparison between actual and desired outputs.

Thus only the strongest activation energies are adjusted; weak signals get progressively weaker unless the magnitudes of their response become comparable to those of the "winners".

Three examples which utilize forms of supervised learning will be discussed here. Perceptrons and Adalines will be discussed since they are the immediate predecessors of backpropagation, which was chosen for use due to its history of success.

2. Perceptrons

The perceptron was developed by Frank Rosenblatt in the late 1950's and early 196 's for use in identifying optical shape patterns and was inspired by the theoretical workings of the human eye. The perceptron is a purely feed forward three layer network wherein only the third layer is involved in the learning process.

The first layer linearizes a two dimensional array of optical inputs and subjects these inputs to an either linear or non-linear transfer function and passes the processed inputs to the second layer via connections of fixed weight. The second layer is utilized for "feature extraction" and compare the inputs from the buffer layer with a threshold value which if exceeded allows further transmittal of the signal to the third layer via another set of fixed connection weights. The third layer, consisting of the actual perceptrons, consists of processing elements that receive

inputs from the second layer feature extractors through variable weight connections and consist of a summer and a step transfer function where the output is zero if the summation of the weighted inputs plus a threshold or bias value of one is less than or equal to zero and is unity if the summation is greater than zero.

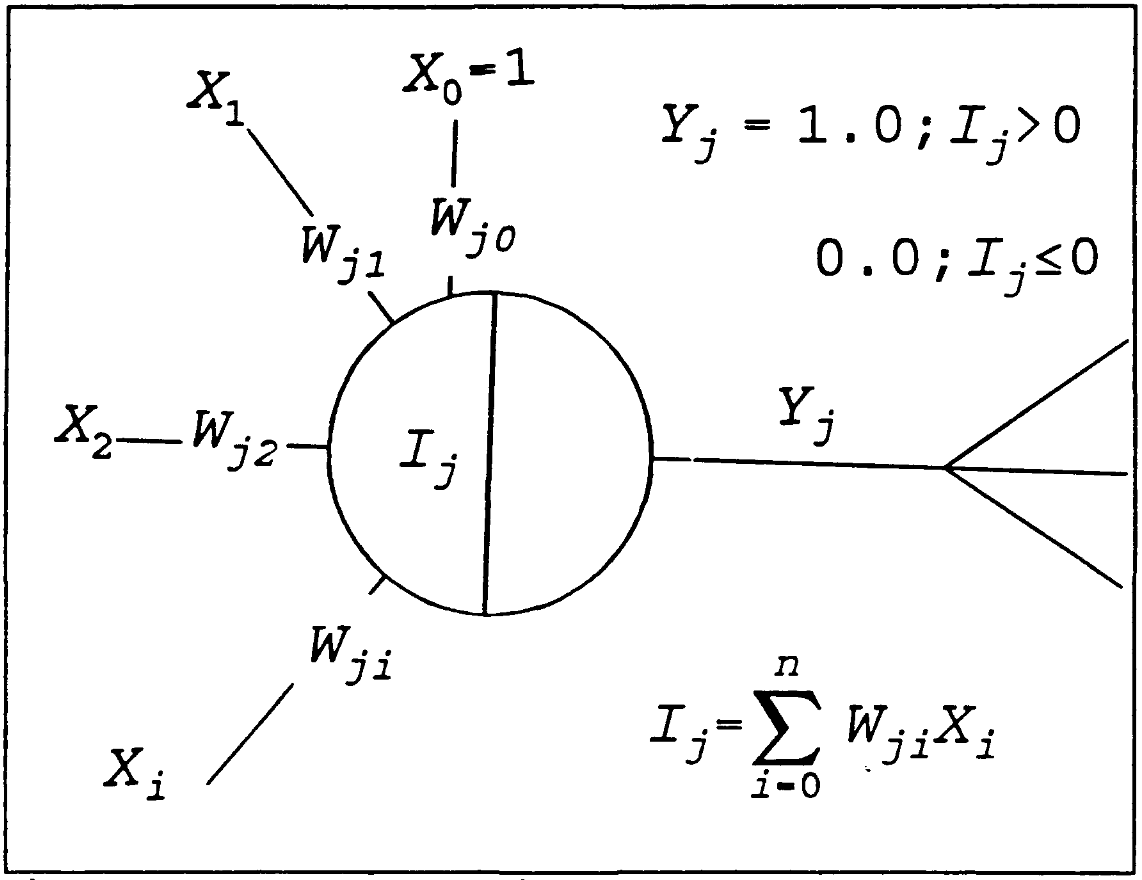

Figure 4 Perceptron Processing Element

Figure 4 shows the binary perceptron processing element. The basic learning algorithm for adjusting the perceptron weights is as follows:

$$I_j = \sum_{i=1}^{n} W_{ji} X_i \quad | \quad Y_j = 1.0,\ if\ I_j > 0; \quad Y_j = 0.0,\ if\ I_j \leq 0 \qquad (2)$$

where Y_j is the actual output , I_j is the summed activation, and W_j is the weighting between the perceptron and the jth feature extractor. In other words, the actual output of the perceptron is compared with a desired output of either zero or one. If they match, all weightings into that perceptron remain as is; if they do not match and the actual output is zero, the weights to that perceptron are incremented a fixed or random amount; if they do not match and the actual output is one, the weights to that perceptron are decremented by that same value.

As mentioned in the previous section, there are two drawbacks to this learning rule. While Rosenblatt proved that the perceptron network would eventually find a set of weights that would place the input vectors into the right categories if that set of weights existed, Minsky and Papert[Ref.14] proved that for this to occur the categories would have to be linearly separable; that is, the solution space of n dimensions would nave to be able to be separated by a hyperplane, or, in multiple perceptron networks, a set of hyperplanes, of n-1 dimensions. They showed that this drawback made it impossible for a single perceptron to solve the exclusive OR problem and implied that this made the perceptron incapable of solving "interesting" problems. The other

drawback was that for multiple perceptrons, there was no real means to determine the direction of weight adjustments in the case of incorrect responses. These problems were later remedied by utilizing multiple layers of processing elements capable of weight adjustment and establishing a feedback loop to help adjust the weights of individual processing elements. Nevertheless, the perceptron was capable of rudimentary shape recognition although it never progressed beyond the experimental stage[Ref.8].

3. Adaline/Madaline

The Adaline or Adaptive linear element was developed by Bernard Widrow and Marcian Hoff[Ref.12] and has a general architecture similar to the perceptron but with some improvements, particularly with respect to determining the direction and magnitude of weight adjustment based on the error in the output. Figure 5 illustrates this architecture.

Like the perceptron, the basic adaline structure consists of three layers. Here, however, it is the middle layer vice the third layer where the learning occurs. In the adaline, the first layer consists of multiple elements which only apply a transfer function to the input value and generate an output of either +1.0 or -1.0. The second layer operates like a classical processing element and performs summation, transfer function, and weight adjustment operations. The third layer consists of processing elements with fixed input weights

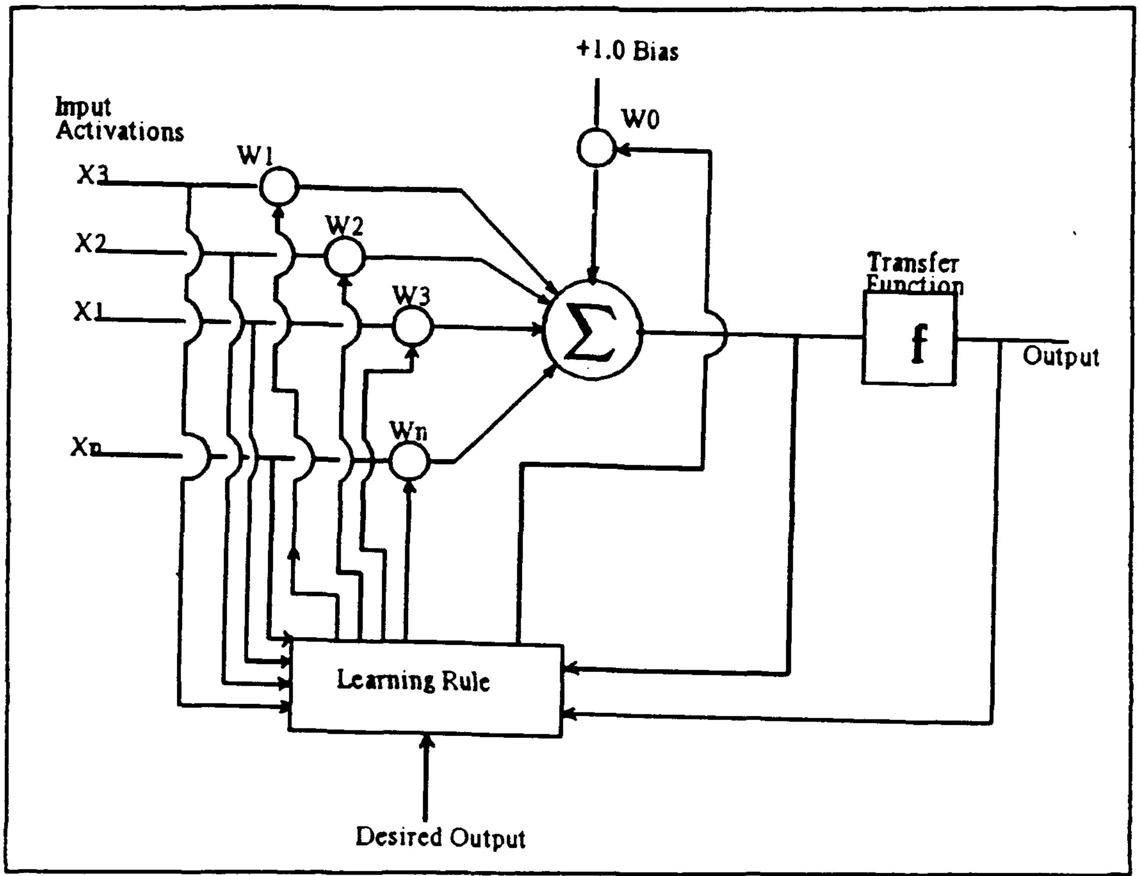

Figure 5 Adaline Processing Element

and performs a linear transfer function on the input.

The middle layer elements, the actual adalines, perform the following operations.

First,

$$I = \sum_{j=1}^{n} W_j X_j \qquad (3)$$

where X_j is the jth input from the previous layer, W_j is its connection weight, and I is the internal activation level. Then,

$$F(I) = SGN(I) \tag{4}$$

where F(I) is the signum function which outputs ± 1.0 depending on the sign of I.

Weights are adjusted by the following algorithm:

$$\delta W_j = \alpha \left(\frac{D_o - I}{N+1} \right) X_j \tag{5}$$

Here D_o is the desired output, α is the learning coefficient, which is values between 0.0 and 1.0, N is the number of weights involved at the processing element, and δ is the increment by which the weight is adjusted. An interesting point about this algorithm is that the weights are adjusted by the difference between the internal activation energy and desired output vice the actual output and the desired output. The effect of this is to permit the weights to continue to be adjusted even after a convergence between actual and desired output is obtained. The effect of the algorithm is to minimize the mean square of the error over the entire set of vectors employed in training.

In summary the adaline has the following advantages over the perceptron. It possesses the means to adjust the weights in the correct direction and with an increment proportional to the existing error. It also continues to adapt even once convergence has been obtained. It is also not

without drawbacks. Like the perceptron, the adaline employs a somewhat linear transfer function and has binary outputs. It also requires the input space to be linearly separable to function successfully. Additionally, if the learning coefficients are too large, and the number of weights exceed the number of unknowns defining the input space, the weights will have the effect of contradicting themselves, thereby preventing convergence of the error function.

The Madaline is a neural network consisting of Many adalines and has first and second layers identical to those of an Adaline network. However, for its third layer, it utilizes a single processing element which is also capable of learning. In essence the Madaline processing element operates to selectively correct the output of the Adalines in the previous level by correcting either the Adaline whose internal activation is farthest in the wrong direction, all of the Adalines operating in the wrong direction, or only the Adaline operating in the wrong direction when the majority of the Adalines are operating in the wrong direction, depending on the particular variety of madaline in use. These Madalines and Adalines have been employed in telecommunications signal processing, non-linear control systems, and in weather prediction[Ref.8].

4. Backpropagation

By far the most successful and popular neural network architecture in use at the present time is back propagation. This architecture addresses all of the drawbacks inherent to the perceptron while still retaining a large portion of the perceptron's basic structure.

a. General Architecture

The architecture still consists of several layers; however, unlike in the cases of perceptrons and adalines, where processing elements capable of learning were confined to one layer, in backpropagation, all layers, that is, input, output, and any number of hidden layers, are capable of having their weights adjusted. Further, the backpropagation network is not confined to three layers; any number of hidden layers are possible. Figure 6 illustrates the backpropagation architecture. The multi-layer learning capability of the backpropagation network allows it to solve non-linearly separable problems, the XOR problem that plagued the perceptron.

b. Processing Element

The backpropagation processing element is similar to both the adaline and perceptron in that it performs three operations: a summing operation, followed by a transfer function, followed by a learning algorithm. A schematic of thie processing element is provided in Figure 7. It differs

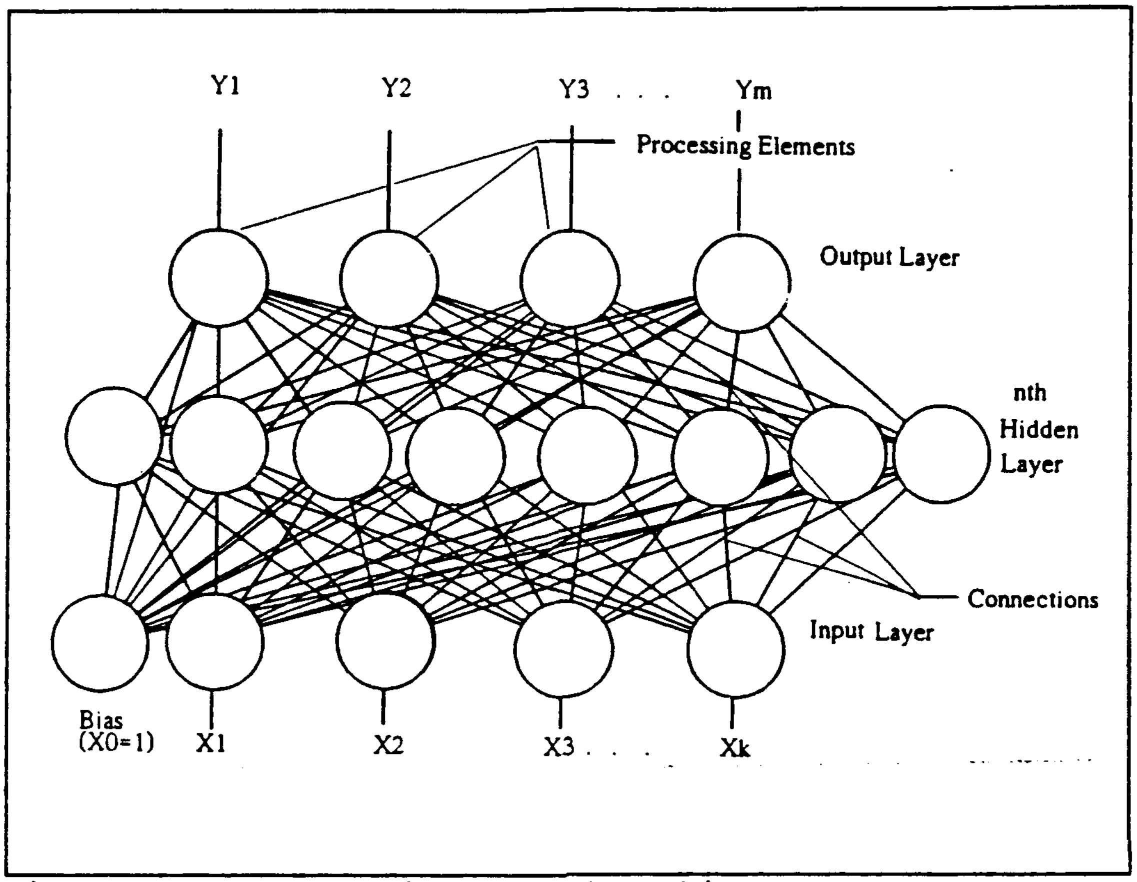

Figure 6 Backpropagation Network Architecture

from the previous processing elements in that it both receives and transmits a non-binary signal.Like the adaline, in addition to the weights associated with the connections between processing elements, there is also a threshold or bias weight associated with each processing element with an adjustable weight but constant input activation of unity.

It also employs a nonlinear transfer function as opposed to a simple binary transfer or linear transfer function in previously discussed networks. This gives the network much greater versatility in mapping the input space and extracting features and makes this architecture

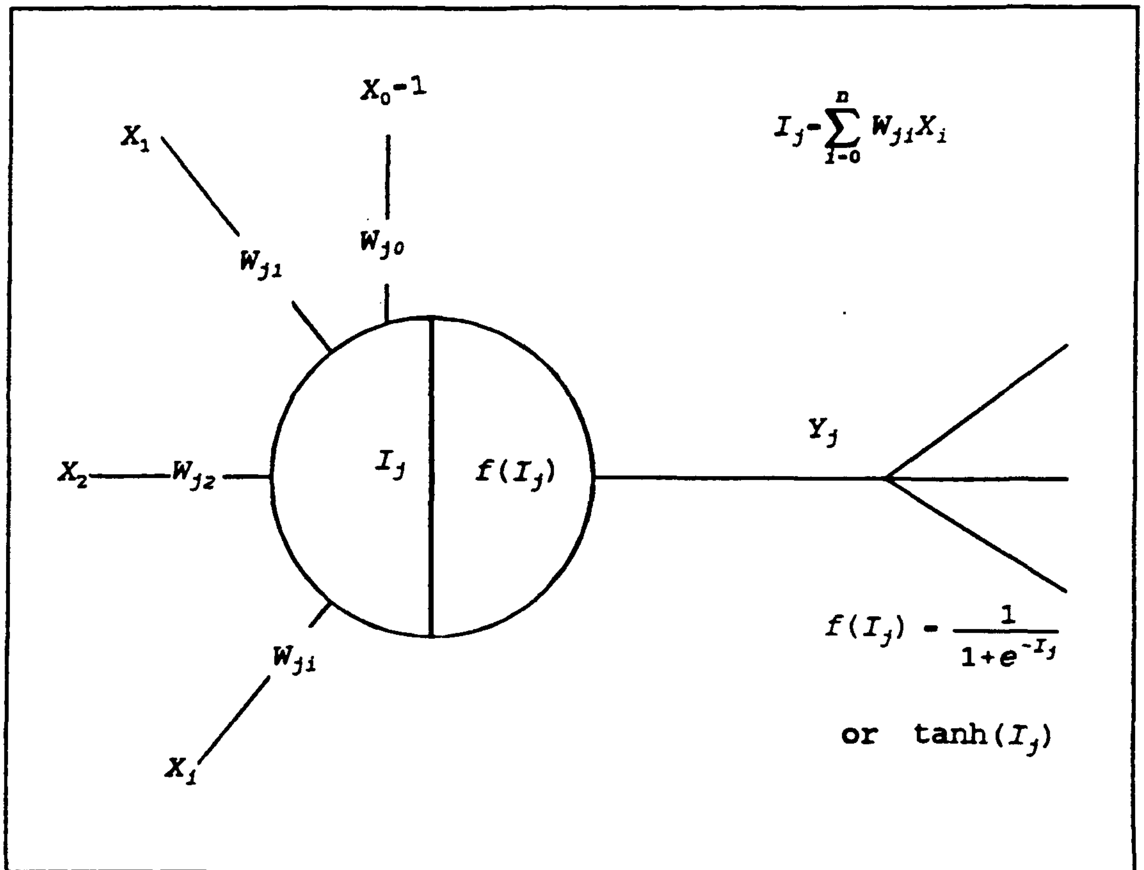

Figure 7 Backpropagation Processing Element

particularly useful in mapping nonlinear relationships. While Rummelhart, Hinton and Williams[Ref.18] indicate that any monotonously increasing transfer function can be employed, the most popular transfer functions currently in use are the sigmoid function, which is defined as:

$$F(I) = \frac{1}{1+e^{-I}} \tag{6}$$

where F(I) is the output of the processing element and I is the summation of all of its inputs. The second most

popular transfer function in use is the hyperbolic tangent, which is defined as:

$$F(I) = \frac{e^{I}-e^{-I}}{e^{I}+e^{-I}} \tag{7}$$

Both of these are employed in the neural networks utilized in this research. These transfer functions are most popular primarily because their derivatives can easily be calculated in terms of the original function, which makes the algorithm more easily programmable. These derivatives are the key to the backpropagation learning rule. A schematic of the common transfer functions is presented in Figure 8.

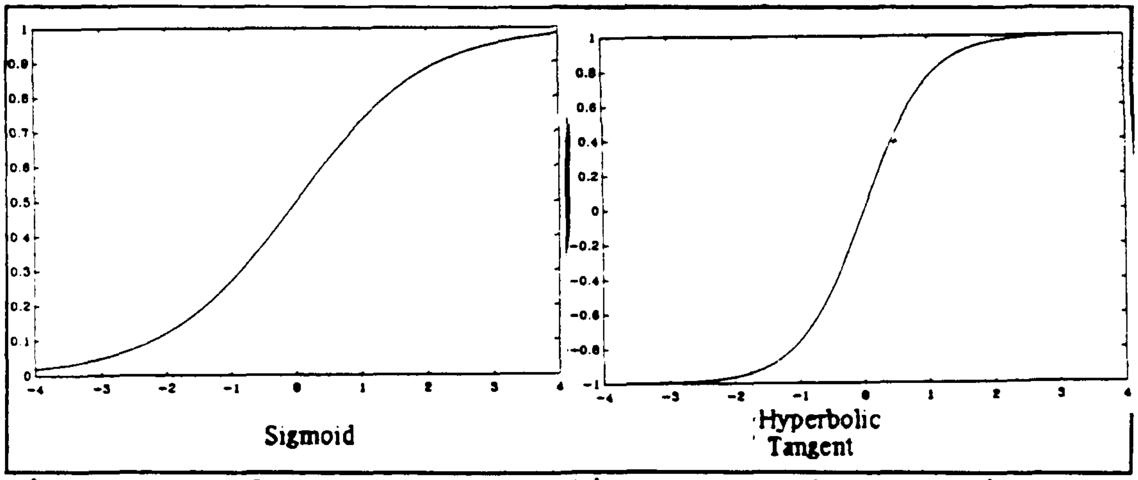

Figure 8 Popular Transfer Functions For Backpropagation

c. *Backpropagation Learning Rule*

The back propagation learning rule is very similar to that used by Widrow and Hoff in the Adaline. As in the case of the Widrow-Hoff rule, the intent of the algorithm is to

adjust the weights is such a way as to follow the path of steepest gradient descent in weight space so as to reach a least mean squares error between the actual and desired output of the network. The means by which this is done, however, is quite different.

Essentially each processing element updates its weights in accordance with the generalized delta rule, which, when neglecting momentum terms, is defined as:

$$\Delta W_{ji} = \alpha (D_{pj} - Y_{pj}) X_{ip} = \alpha \delta_{pj} X_{ip} \qquad (8)$$

where ΔW_{ji} is the change to the connection weight between the jth processing element and the layer in question and the ith processing element in the previous layer; α is a learning coefficient, usually between 0 and 1; D_{pj} is the desired output of the jth processing element upon presentation of the pth training vector and Y_{pj} is the actual output; and X_{pi} is the weighted input from the ith element in the previous layer. To prove that this rule approximates an adjustment of the weights along the gradient of steepest descent in weight space, let E_p represent the overall error found in the network upon presentation of the sample vector p.

$$E_p = \frac{1}{2} \sum_{j=1}^{n} (D_{pj} - Y_{pj})^2 \qquad (9)$$

The object is to prove:

$$-\frac{\partial E_p}{\partial W_{ji}} = \delta_{pj} X_i \tag{10}$$

Using the chain rule,

$$\frac{\partial E_p}{\partial W_{ji}} = \frac{\partial E_{pj} \partial Y_{pj}}{\partial Y_{pj} \partial W_{ji}} \tag{11}$$

$$\frac{\partial E_p}{\partial Y_{pj}} = -(D_{pj} - Y_{pj}) = -\delta_{pj} \tag{12}$$

$$Y_{pj} = W_{ji} X_{pj} \tag{13}$$

Thus,

$$\frac{\partial Y_{pj}}{\partial W_{ji}} = X_{pj} \tag{14}$$

Substituting (14) into (11) yields:

$$-\frac{\partial E_p}{\partial W_{ji}} = \delta_{pj} X_{pj} \tag{15}$$

Since,

$$\frac{\partial E}{\partial W_i} = \sum_{p=1}^{m} \frac{\partial E_p}{\partial W_{ji}} \tag{16}$$

,the change in W_{ji} approaches being proportional to the gradient descent in weight space when minimizing the overall error. If there was no change in the weighting, then this would be exactly so but since the weights change at each presentation, the rule only approximates the path of steepest descent. Fortunately, if the change in weights is kept small between presentations of input vectors, the approximation approaches the exact path.

Rummelhart extends this proof to processing elements with nonlinear transfer functions. The only real difference is that with nonlinear transfer functions, the derivative of the transfer function has to be calculated. Here,

$$\Delta_p W_{ji} = \alpha \delta_{pj} Y_{ip} \tag{17}$$

where,

$$\delta_{pj} = (D_{pj} - Y_{pj}) F'(I_{pj}) \tag{18}$$

where, F' is the first derivative of the transfer function and I_{pj} is the summation of weighted outputs, X_{ip}'s, from the previous layer. For processing elements in intermediate layers where there is no desired output available for computation of the error, the error is determined by feeding back the weighted errors from the processing elements from the next layer. In other words, for the ith element in the (k-1)th intermediate layer, the error term is backpropagated from all of the jth elements from the kth layer as follows:

$$\delta_{ip} = F'(I_{ip}) \sum_{j=1}^{n} \delta_{pj} W_{ji} \tag{19}$$

Thus the operation of the network is as follows. First the input vector is presented to the input layer and transmitted through each successive layer up through the output layer. The actual outputs are compared with the desired outputs and error signals are computed in accordance with the Generalized Delta Rule, Equation (17), and then adjusting the weights leading to the output layer. The errors computed in the output layer are then used to compute the error in the previous layer processing elements in accordance with equation (19) and adjusting the weights leading to that layer accordingly. This process continues backwards through the

network until the weights leading to the input layer are adjusted. Then the next vector presentation occurs.[Ref.18]

d. Practical Considerations and Modifications

Although the backpropagation algorithm is quite robust and has proven itself capable of solving a wide variety of problems, its use is not without its drawbacks. As experience in using backpropagation has grown, a number of embellishments and modifications have been developed to resolve practical difficulties inherent to the backpropagation algorithm. In this section a number of practical considerations and means to overcome them will be discussed.

(1) Limitations of Transfer Functions. While the utilization of non-linear transfer functions is the source of a great deal of power in the backpropagation algorithm, it is also the source of a few drawbacks. A quick view of the sigmoid and hyperbolic tangent functions will reveal that the functions asymptotically approach 0.0 and 1.0, or -1.0 and +1.0 , respectively. This means that there will always be an error associated if the desired outputs are at these asymptotes. Rummelhart[Ref.18] recommends that, to improve the chances of convergence, or minimization of the error, or at least to reduce computation time, one should set these types of desired outputs to, for example, 0.1 and 0.9 instead of 0.0 and 1.0 Another alternative is to reduce the standards of convergence, taking the impossibility of a complete

convergence into consideration. At these asymptotes it is also readily noticeable that the derivatives of the transfer function approach zero. Thus if the activation energies get very large in either a positive or negative sense, the derivatives approach zero and no learning takes place. This is generally caused by allowing the absolute value of the connection weights to become excessively large and is called saturation. Scott Fahlmann[Ref.22] indicates that this can be alleviated to some extent by introducing a small positive number to the derivative. Another possible remedy is to limit the size of the delta weights by reducing the learning coefficient, α [Ref.8]. This increases the number of iterations required for the weights to transit from zero to the very high value weights. There is thus a greater possibility of attaining convergence before saturation sets in.

(2) Initialization of Connection Weights. In the original backpropagation networks, all connection weights were initialized with values of zero, and all weight adjustments were made by the delta rule. This resulted in symmetric weight adjustments for all connection weights feeding into each individual processing element due to the proportionality of weight adjustments to the propagated error inherent to the delta learning rule. While there were a number of problems that could be solved with this arrangement, many more mappings

requiring assymmetric weights could not be learned. This problem can be readily overcome by distributing the weights randomly about small values around zero. In this manner all weights start out at different initial values and the pattern of symmetry can be broken out of from the start. Most backpropagation programs currently in use employ this randomization scheme.[Ref.18]

(3) Learning Coefficients. A critical determinant of the size of the weight changes from one vector presentation to the next, along with the magnitude of the error function is the value of the learning coefficient. If the learning coefficient, α. If α is large, there is a tendency for the weights to fluctuate wildly, increasing the probability that the weightings will not be able to home in on local or absolute minima in weight space, especially if the minimum is deep and narrow in the weight space. Smaller learning coefficients allow the network to sense the contour of the weight space more accurately, thereby reducing the probability that a deep narrow minimum would be missed. The drawback of the low learning coefficient is that if it is too low, the weight adjustment will be excessively slow and convergence time will be extended as a result. Rummelhart, Hinton, and Williams[Ref.18] recommend a learning coefficient of between 0 and 2 for most applications; Neuralware Incorporated advocates a learning coefficient from between 0 and 1.

Further, they recommend that the learning coefficients be reduced in value as learning progresses so as to allow rapid exploration of the weight space during the initial learning followed by increasingly finely tuned adjustments as learning progresses. Additionally, practical experience indicates that as one increases the number of processing elements in a network, the learning coefficient should be reduced[Ref.23].

(4) Modifications to the Delta Learning Rule. In an effort to improve the speed and efficiency of the basic Delta Rule, a number of modifications have been suggested. A major problem in basic delta learning is the tendency of the algorithm to get locked into small variations of the error surface in weight space. While the use of small weight changes reduces the network's tendency to "fibrillate", where the weights and errors fluctuate wildly with minimal net reduction in the error function, it seems to increase the network's vulnerability to these shallow valleys in the error surface. A simple means to escape these valleys once entrapped is to change all the weights by a fixed amount and resume learning from that point. Neuralware's Professional II neural network simulator provides for this in its jog weights function.

Modifications to the basic learning algorithm that reduce the vulnerability to this problem include the inclusion of a momentum term and utilization of a cumulative error function. The inclusion of a momentum term in the delta

rule has the effect of increasing the motion of the weights in the direction of steepest gradient descent by reinforcing the change in weights in the current vector presentation with a factor based on the change of weights due to the previous vector presentation. Here the basic delta rule is altered to:

$$\Delta W_{jip} = \alpha_{jp} X_{ip} + \beta \Delta W_{ji(p-1)} \qquad (20)$$

where ß is the momentum factor, and p and p-1 refer to the current and previous presentation, respectively. This has the effect of filtering out the high frequency variations in the error surface.

In the cumulative delta rule, the weights are not immediately adjusted after each vector presentation. Rather, the errors are accumulated over the entire or partial set of training vectors, called an epoch, and the weights are then adjusted. This has the effect of adjusting the weights to minimize the global error function as opposed to the error of each individual vector. While this greatly reduces the network's tendency to fibrillate, it also tends to increase the learning time, as the weights are only updated once each epoch[Refs.8 and 23]. Nevertheless, the response to the global error inherent to this modification is increasingly important as the complexity of the solution space increases and thus is used extensively in this research.

5. Unsupervised Learning: An Example

Because unsupervised learning has several inherent advantages over supervised learning, namely independence from an extensive data base, it shows great promise in machinery diagnostics applications and, although it is not employed in this research, warrants some discussion. An excellent example of this genre of neural networks is Binary Adaptive Resonance Theory, (ART1), developed by Steven Grossberg[Ref.24].

The network utilizes two layers of processing elements interconnected by a series of connections called long term memory. The lower layer of vectors performs transfer functions on an input vector and transmits an activation signal to the second layer via the long term memory connections.

The upper layer utilizes a competitive learning algorithm and all second layer processing elements currently possessing reference vectors compete until only one of these processing elements remains active. The winning processing element then transmits a signal related to its reference vector to the lower level and creates a new activation signal.

This activation signal is then compared with the activation signal associated with the original input vector and a magnitude of the error between the two is calculated. If this error value exceeds a threshold, the upper level processing element generating the new activation signal is removed from the competition and the other upper level processing elements possessing reference vectors continue

competition until there is another winner. It then transmits a new activation signal to the lower layer and comparison of the error is compared once again with the threshold.

This process continues until a winning upper level processing element is able to generate an activation signal within the error threshold. If no such processing element is located, a new processing element is brought on line with a reference vector related to the original input vector. If a winner is found within the threshold criterion, the original input vector is incorporated into that processing element's reference vector.[Ref.6]

This scheme has several inherent advantages. First it acts as a pattern classifier and does not require the desired output vector associated with supervised learning to function. Second, it is capable of placing new patterns outside its threshold limitations into new categories. Its drawback is that this particular algorithm is only capable of handling binary inputs; however, Grossberg has developed other algorithms with greater versatility and is working on a non-binary version, ART3, which is still in the developmental stage.

6. Why Neural Networks?

Neural Networks possess several traits that make them an attractive alternative to conventionally configured expert systems. First, many are capable of discerning non-linear

relationships. Second, they are capable of functioning with a certain degree of background noise and erroneous information with minimal degradation of their pattern recognition abilities. Third, they have the ability to generalize, having the ability to classify previously unseen vector patterns into existing and in some cases new output categories. They are also capable of identifying multiple faults. These are all areas where traditional expert systems typically fall short. Moreover, neural networks are data based rather than rule based. This means that they may be capable of correctly discerning relationships previously hidden from the best of "experts".

Neural Networks are not without their disadvantages. They, like all computers, are capable only of manipulating numbers and require an engineer to discern the intelligence of their output. Their success is largely limited to the quality of the data that they are provided. If the input vectors provided are inadequate to describe the decision space fully, then their likelihood for success is small. Again, they require an engineer to provide the proper inputs. Finally, they may be able to discern new relationships, but the relationships themselves remain hidden; all that is seen external to the network are the input and the output vectors. It is generally believed that the relationships are somehow hidden in the connection weights and the hidden layers but meaningful extraction of this information has yet to occur.

The question might be asked whether a neural network should theoretically be capable of recognizing patterns in vibration signatures. Kolmogorov's Theorem indicates that any continuous function can be represented exactly by a 3 layer neural network with n input nodes, 2n+1 hidden nodes, and m output nodes, and presumably mechanical systems can be at least approximated by at least piecewise continuous functions. Therefore, at least theoretically, the neural network should be able to succeed. Unfortunately, nobody has yet been able to develop a Kolmogorov neural network. Nevertheless, backpropagation does possess a number of the features identified by Kolmogorov.[Ref.19]

Neural networks would appear to have potential in numerous fields, including machinery diagnostics. It is the task of this research to determine whether this potential can be realized in the region of machinery diagnostics. In order to accomplish this it will be necessary to demonstrate the validity of the claims made above while overcoming the limitations also duly cited. In order to accomplish both a good basis in machinery diagnostics theory is required.

III. MACHINERY DIAGNOSTICS OVERVIEW

Vibration analysis is among the most powerful tools available for the detection and isolation of incipient faults in mechanical systems. Among the methods of vibration analysis in use today and under continuous study are broad band vibration monitoring, time domain analysis, and frequency analysis. All have varying degrees of utility in machinery condition monitoring and diagnostics and have characteristics that lend themselves particularly well to specific applications. Since the effectiveness of a neural network is directly related to how effectively the chosen inputs define a particular decision space, the selection of the optimum vibration parameters for inputs to the neural network is critical. Thus a good understanding of elementary machinery diagnostics techniques is essential.

A. SOURCES OF VIBRATION

In mechanical systems any mechanical component which periodically comes in contact with a second component to transmit an axial, radial or torsional load is a potential source of mechanical vibration. In a machine with a gear train the principal components involved with load transfer will be its torsional power source, such as a motor, the gear meshes, the bearings, and those items that interconnect them, the

shafts. Additionally, because vibrational isolation is seldom complete, additional extraneous sources of vibration will also be present. The diagnostician is generally interested in extracting the vibrations created by specific machinery components and ignoring the other sources as extraneous noise. In this study we are particularly interested in the vibrations generated by the rotating machinery's gears, bearings, and shafts. As such, the discussion will be limited to these sources of vibration.

1. Gear Vibration

In a gear train, the gear mesh is the dominating source of mechanical vibration. This vibration primarily stems from the nonuniformity of the transmission of angular motion from one gear to its mate. The nonuniformity of the angular motion occurs due to geometric deviations of the contact surfaces from the ideal involute shape and the elastic deformation that any mechanical system undergoes when transmitting a load[Ref.25]. The geometric deviations are in turn caused by profile and pitch errors, and variations in the surface finish of the teeth. Tooth impacts, oil and air ejection as these fluids are forced across the contact surfaces also contribute. Finally, torque fluctuations and deflections of the gear box can also be sources of vibration in gears. Clearly, any damage that occurs to the gear contact

surface as well as other mechanical linkages to the gear mesh will also have an effect on the gear's vibrations[Ref.26].

These factors generally contribute to excitation at the gear mesh frequency and at the sidebands associated with the offending gear. The gear mesh frequency is obtained from the frequency of impacts between the teeth of each gear and is calculated by the equation,

$$F_{gm} = N_t F_s \tag{21}$$

where F_{gm} is the gear mesh frequency, F_s is the shaft rotational frequency, and N_t is the number of gear teeth. Regardless of damage present, this signal and its harmonics is always present. The sidebands are caused by the frequency modulation of the gear meshing due to backlash, eccentricity, loading, bottoming, and impacts caused by defects or damage to the gear. These sidebands generally differ from the gear mesh frequency by the rotative frequency of the affected gear and its harmonics[Ref.27]. The magnitude of these sidebands tends to increase as damage occurs to the gear.

Randall[Ref.28] indicates that a majority of gear faults can be identified using the frequencies about the first three harmonics of the gear mesh frequency. Further, while impact faults can be readily detected at these frequencies, Favaloro[Ref.29] states that even wear over all of the teeth is very difficult to detect until the most advanced stages of

damage. Because of this most gear faults to be studied in this research will be due to damage to a single tooth.

2. Bearings

Bearing vibrations occur for much the same reasons as gears. However, because bearings are not situated directly along the power transmission train and support largely static loads, they characteristically generate a small vibration signal until the damage inflicted upon them reaches advanced stages. Because of the low magnitude of these signals, they are often masked by much stronger gear related signals. Partially because of this belated detection of trouble, antifriction bearings are among the most common causes of machinery failure in moderately sized machines.

The frequencies associated with bearing related signals generally depend on the location of the damage, the dimensions of the bearings, and the shaft rotation speed. In general fundamental bearing related frequencies can be obtained by calculating the impact frequency for a ball in the bearing impacting a fault on the inner or outer race and the impact frequency for a fault located on the ball impacting other bearing components. These impact frequencies adhere to the following formulae:

$$F_{bo} = \left(\frac{N_b}{2}\right)(F_s)\left(1 - \frac{PD}{BD}COS\phi\right) \qquad (22)$$

$$F_{bi} = \left(\frac{N_b}{2}\right)(F_s)\left(1+\frac{BD}{PD}COS\Phi\right) \tag{23}$$

$$F_{bb} = \left(\frac{PD}{2BD}\right)(F_s)\left(1-\left(\frac{BD}{PD}\right)^2 COS^2\Phi\right) \tag{24}$$

where F_{bo} is the outer race impact frequency, F_{bi} is the inner race impact frequency, F_{bb} is the ball impact frequency, N_b is the number of balls, F_s is the shaft rotative frequency, PD is the pitch diameter, BD is the ball diameter, and φ is the contact angle between the ball and inner or outer race. These formulae reflect the fact that the balls must travel along the races at a speed that is the average of the relative tangential speeds of the inner and outer races and the fact that, because of the smaller diameter at the inner race, the balls must impact a defect on this race at a higher frequency than a fault on the outer race.[Ref.27]

While in the low frequency region the calculation of these frequencies is relatively straightforward, there is also a tendency for other vibration sources to dominate. Because of this, many sources recommend that higher frequencies be used to find bearing signatures. Sandy[Ref.30] recommends that the region of between one to seven times the inner race impact frequency be monitored for bearing signals while Collacott [Ref.31] reports that while 80 percent of bearing faults demonstrate symptoms at one to two times the impact frequency,

20 percent manifest themselves at "very high frequencies". Sandy also indicates that bearing faults can manifest themselves at frequencies as high as 5 to 35 kHz.

3. Shafts

Shafts generally produce vibration signals at their rotational frequency and its harmonics. Shafts are also prone to a number of different faults, all of which register at the shaft rotative frequency. In the case of bent shafts and shaft misalignments, the second harmonic is the dominant frequency in 90 percent of the cases[Ref.31]. Imbalances in the shaft or load characteristically generate a dominant signal at the shaft rotative frequency but there tends to be a phase shift as well. Mechanical looseness can also introduce increases in the shaft rotational frequency but also characteristically involves higher harmonics as well[Ref.27].

4. Extraneous Signals

Intertwined with the relevant signals that can provide the troubleshooter with valuable information are a number of undesirable signals from countless other sources. Characteristically they include electro-magnetic signals from nearby induction motors and other electrical power supplies as well as vibrations emanating from other machinery in proximity. Electro-magnetic signals generally occur at multiples of the power generation frequency and are usually quite stable, thereby proving fairly easy to identify[Ref.27].

The other extraneous signals can often be averaged out of the signal being monitored by utilizing a time synchronous averaging technique. In this technique, a trigger signal is transmitted to the monitoring device from a proximeter that is monitoring the machine in question. The trigger signal then causes the signal analyzer to take sample measurements for averaging only at the synchronous speed of the machine being monitored. This causes the asynchronous signals to average out to zero as the number of averages gets large. As an alternative, asynchronous averaging can also be used to minimize the influence of extraneous noise on the vibration signal under investigation.[Refs.27 and 32]

Another extraneous source of difficulty when attempting to monitor a given machine is the tendency for that machine to change speed from time to time. This generates confusion in the analysis of vibration signals by shifting the frequencies associated with various components up or down by a factor of some multiple of the change in frequency. For example, if the rotational speed changed from 30 to 31 Hz and one was interested in a gear mesh frequency for a 15 tooth gear that is nominally located at 900 Hz, that signal will change to 915 Hz. If this effect is not taken into account, it is very easy to misidentify signals. This can be automated away by utilizing an external trigger source that measures speed of the machine being monitored and a feature found on most dynamic signal analyzers called ordering. When activated

this feature normalizes all frequencies in terms of the operating frequency of the machine being monitored. This has the effect of holding the relative positions of the various frequencies constant so that more trouble free analysis can take place[Ref.27]. If an external trigger source is not available, then the frequency shift must be taken into account mentally or by hand.

B. MACHINERY MONITORING TECHNIQUES

Vibration signals are essentially measurements of a mechanical system's total dynamic response to all forms of internal and external excitation acting on the system at a given time. These measurements can be made using displacement, velocity, or acceleration transducers. While all of these measurements have their place in machinery condition monitoring, the most popular at present involves acceleration measurements. These measurements can then be represented in three ways. The most direct method is to simply measure the overall level of vibration. However, these measurements tend to downplay the dynamic nature of the excitation. The least complicated way to incorporate this is to plot these responses with respect to time. Another method is to plot these responses with respect to frequency. This section explores some of the techniques used to extract pertinent information using each of these representations of the vibration signal.

1. Broad Band Monitoring of the Overall Vibration Level

Broad band overall level monitoring provides a broad level of vibration occurring at a measurement point. This simple approach is often used for day to day trending of the relative health of a machine. The setup usually involves a velocity or acceleration transducer and a vibration meter which provides an RMS vibration level over a broad frequency range, thereby being capable of receiving excitation along a large range of frequencies. While useful in detecting a fault, it is virtually useless in diagnostics because of the lack of frequency information. Its capability in fault detection is also limited since it tends to be most strongly influenced by the dominant frequencies chatracteristic of the machine. If a fault occurs on a component not associated with a dominant frequency, the fault will not be detected until the damage reaches an advanced stage. However, this method lends itself to easily portable equipment, is inexpensive, and requires no special training to use.[Ref.31]

2. Time Domain Vibration Monitoring

A large number of techniques are available that manipulate the time domain signature of machinery vibrations. Among these are waveform analysis, index analysis, time synchronous averaging, and the analysis of statistical parameters. In a broad band mode these techniques can prove very useful in detecting machinery faults. By using filtering

techniques, and narrowing the bandwidth, characteristic frequencies can be isolated and monitored to provide a useful diagnostic tool.

a. *Waveform Analysis*

Waveform analysis involves the study of the time-amplitude plot of the vibration signature. It can be used to determine the degree of randomness in a signal as well as identify periodicities. Damage affecting a particular locality on a machinery component can often be identified, especially after the fault has gone beyond the incipient stage. An example of a machinery fault in a time domain plot is presented in Figure 9. Waveform analysis can also be used to identify beats and vibrations not synchronous with shaft rotation which are often averaged out in techniques such as synchronous averaging[Ref.32].

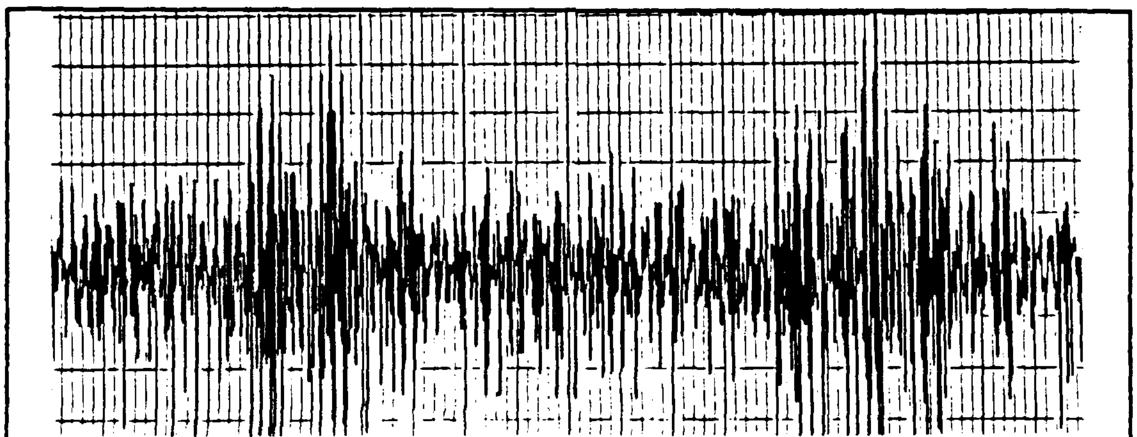

Figure 9 Time Signal for Bent Shaft Fault

b. Time Domain Indexing

In many condition monitoring programs, it is highly desirable to reduce the the amount of data recorded to the minimum required to get the job done. As a result, indexing in both the time and frequency domain are quite popular. Three indexing parameters are most common. The first is peak level, which is merely the maximum value of the vibration over a given time span. Because it only takes one spuriously high reading to possibly indicate a fault condition, it is not considered very reliable. The most commonly used index is RMS level which is statistically based and can provide fairly good results. However, as mentioned in the broad band monitoring section, RMS averaging usually results in masking out the smaller signals which may be significant. Often, especially in its earliest stages, a fault condition will manifest itself through vibration measurements occasionally rising above the RMS level but not often enough to significantly affect it. To provide an indication of both peak and RMS values a third parameter known as crest was developed. This value is simply the difference between peak and RMS values. In many incipient faults, this value will increase at first and then, as the damage builds and RMS level catches up to the peak values, it will decrease. If a time record is kept such a fault would be detected; if not, such a fault indication could easily be missed.[Ref.32]

c. *Time Synchronous Averaging*

Time synchronous averaging involves averaging a signal over a large number of cycles synchronous with the rotational speed, thus having the effect of eliminating extraneous vibrations from other machinery components. It is often used in diagnosing faults in multiple gear trains to mask out adjacent gear vibration as well as in other areas where extraneous noise is high.[Refs.27 and 32]

d. *Statistical Analysis*

A number of statistical parameters which have been extracted from time domain signals have proven particularly capable in detecting incipient faults in machinery components. Among these are included the probability density function, probability distribution function, and several higher moments of the probability distribution function.

The probability density function is defined as the length of time that a signal occurs at a certain amplitude normalized by the length of the time record over which the samples are taken. The equation for this is:

$$p(x \le X(t) \le x+\Delta x) = \sum_{i=1}^{n} \frac{\Delta t_i}{T} \qquad (25)$$

where $X(t)$ is a vibration signal, x is a certain amplitude, Δx is an incremental amplitude, Δt_i is an incremental time window, and T is the time record length. By

monitoring the shape of this curve, which for a normal machinery component takes on the Gaussian bell shape and tends to widen at the extreme amplitudes with a corresponding drop at the mean amplitudes as damage occurs, incipient faults can be detected.

The probability distribution function is determined by integrating the probability density function over all time. This function enhances the density function's characteristic broadening at the extreme amplitudes when damage occurs and hence can enhance detection of the fault.

The moments of the probability distribution function follow the general form:

$$M_n = \int_{-\infty}^{\infty} x^n p(x)\,dx; n=1,2,3,\ldots \tag{26}$$

The first and second moments of the probability density function are the arithmetic mean and mean square values, used heavily in this research. The more popular of the higher moments include the third moment or skewness which when the mean is subtracted and it is normalized with respect to the standard deviation, takes on the form:

$$\sqrt{\beta_1} = \frac{\int_{-\infty}^{\infty} (x-\overline{x})^3 p(x)\,dx}{\sigma^3} \tag{27}$$

Another popular moment is the fourth moment or kurtosis, which takes on the form,

$$\sqrt{\beta_2} = \frac{\int_{-\infty}^{\infty} (x-\bar{x})^4 p(x)\,dx}{\sigma^4} \tag{28}$$

In general, the odd numbered moments indicate the peakedness of the signal while the even numbered moments yield indications of the spread of the amplitudes[Ref. 33].

In fault detection the odd numbered moments are usually around zero whereas the even numbered moments react strongly when confronted with impact type damage. Thus the more useful fault detection moments are the even moments. Kurtosis is considered the more useful than the other even moments. Kurtosis tends to strike a balance between the mean square or variance, which are somewhat insensitive to incipient faults, while higher moments are overly sensitive.

The benchmark for kurtosis is based on its value relative to that existing for a Gaussian distribution, where kurtosis is 3.0. If the kurtosis is greater than 3.0 then damage is probably occurring. Further the location of the kurtosis greater than 3.0 in the frequency spectrum is significant, with the higher frequency an indication of greater damage.[Ref.33]

All of these time domain signals and parameters have their uses; however, with the possible exception of the

raw time signal and the connection between the standard deviation of the amplitude to fault severity, these parameters are most valuable in the early detection of faults and not so much with the diagnosis of its location. By far the most convenient method by which to locate machinery faults associated with certain frequencies is through exploitation of the frequency domain.

3. Frequency Domain Vibration Analysis Techniques

Mathematically, the primary method of obtaining a frequency domain plot involves taking the Fourier Transform of the time signal:

$$F(\omega) = \frac{1}{2}\pi\int_{-\infty}^{\infty} f(t)\,e^{-j\omega t}dt \tag{29}$$

Until fairly recently most analysis of the frequency domain was extremely time consuming because of the calculation of the Fourier Transform of the vibration signal was computationally prohibitive. At this time there was no recource but to use digital filters to sweep the frequency spectrum to obtain frequency domain information. With the advent of the Fast Fourier Transform(FFT), however, the frequency spectrum has become easily accessible and is currently the most popular mode of vibration analysis [Ref.31].

a. *Linear Spectrum*

The most direct frequency analysis can be accomplished by observing the linear frequency spectrum, which is obtained by performing an FFT directly to a time signal. Its equation in continuous form is identical to that of Equation (29).

These plots can be modified to present more elaborate information if they are arranged in a cascade plot, which plots a series of time consecutive linear specrtra in a three dimensions. This can prove useful when analyzing machines undergoing transient conditions but the time intervals between the plots becomes limited by the required size of the time record, which varies inversely with the frequency span of interest.

In steady state conditions the cascade plot has the tendency to become excessively cluttered. A variant of this involve plotting the average of a series of linear spectra. This tends to mask out spurrious noise and is used to great extent in this research, where 15 time averages per measurement were used. Other variants include using the indicies mentioned in the previous section and using a masking algorithm which subtracts the baseline from the raw frequency spectrum[Ref.32].

b. *Power Spectrum*

The power spectrum is similar to a linear spectrum except here the discrete elements of the fourier transform are squared. The continuous form equation for this parameter is:

$$G_{xx} = \frac{2}{T}\int_0^{\frac{T}{2}} |F(j\omega)|^2 d\omega \tag{30}$$

where G_{xx} is the power spectrum, T is the period, $F(j\omega)$ is the frequency domain representation of the function, and ω is the angular frequency. This representation is a more direct representation of the power distribution of the signal, hence its name[Ref.31]. In general because of the squared nature of this representation, peaks are more strongly accentuated than in the linear spectrum. Conversely, valeys are lower as well, making low value excitations as might be expected from small lightly loaded machinery components even more difficult to measure.

c. *Cepstrum*

Originally the Cepstrum was defined as the power spectrum of the of the logarithm of the power spectrum, but, in order for it to appear more similar to the autocorrelation function, it was later altered to the inverse Fourier Transform of the logarithm of the power spectrum or:

$$C(\tau) = \mathscr{F}^{-1}[\log\{G_{xx}(j\omega)\}] \tag{31}$$

This parameter has the effect of compressing the frequency spectrum into families of frequencies of the same frequency spacing. Thus harmonic frequencies generally compress into a single "quefrency", as do sidebands. These parameters have certain advantages over individual sideband analysis. First, they are more easily detectable as individual sideband may be masked while the cepstrum, representing the entire family of sidebands is not. Second, in the diagnostics of multiple gear mesh and bearing machines it is often difficult to discern between two different sidebands of similar frequency modulation. This is exacerbated by the tendency for the sidebands to change their modulation slightly from one sideband to the next. This makes identification of the sideband's origins difficult in some cases. With cepstral analysis the frequency spacings that tend to float in the frequency domain are averaged over the entire family of frequencies. Hence its source is more easily identifiable. Thirdly the cepstrum has a tendency to normalize its amplitude, thereby making it much less susceptible to extraneous vibrations. In this research, the cepstrum decibel(dB) level variation over a series of tests remained small whereas the changes from sideband to sideband could be commonly as large as 8.0 dB[Ref.35]. While sideband analysis appears to be one forte for the cepstrum, it has also been noted to be very successful in identifying bearing related faults as well, being documented as the principle indicator

for bearing faults in at least one rule based diagnostic system[Ref.36].

IV. A SIMPLE MACHINERY DIAGNOSTICS MODEL

In order to explore the behavior of backpropagation neural networks in a machinery diagnostics environment a series of experiments were conducted using very simple machinery diagnostics models. The purpose of these experiments was to determine whether the application of neural networks in machinery diagnostics warranted further study. In addition, it was intended to utilize a series of these simple diagnostics models as the basis for the more complicated follow-on machinery diagnostics systems to be discussed in detail in Chapter VI.

A. PROBLEM FORMULATION AND MODEL DESCRIPTION

In these experiments a simple diagnostic model was established based on current practice in machinery condition monitoring programs aboard U.S. Navy surface ships. In these programs, vibration data is obtained periodically by condition monitoring teams, who then send the data ashore for analysis. During the analysis an extensive data base is accessed and the current readings are compared to an established baseline and a magnitude difference in decibels(dB) is obtained. In the current Navy program, a general fault condition is deemed to exist when the current amplitude exceeds the baseline by more than 6.0 dB, barring experientially based dB differences to

the contrary.[Ref.36] The model used for the preliminary experiments monitored four discrete frequencies each associated with a separate machinery components in a hypothetical rotating machine. Amplitude readings would be taken at each of the associated frequencies and compared to a baseline. The absolute value of these dB differences were then entered as a single four dimensional vector into a neural network consisting of four input PE's, any number of hidden PE's, and four output PE's. The output required was a severity indication for each of the inputs based on the rules cited in Table I.

Table I Simple Model Severity Criteria

dB Difference	Network Desired Output	Nomenclature
0.0 - 2.5 dB	0.0	No Fault
2.5 - 4.0 dB	0.3	Low Severity
4.0 - 6.0 dB	0.6	Moderate Severity
6.0 + dB	0.9	High Severity

These severity levels would be associated with a specific course of action to be taken by the operator. For example, if a low severity indication was received it might warrant more frequent observation; if a moderate severity level was indicated, it might warrant replacement at the next scheduled

maintenance period; if a high severity level registered, immediate replacement might be warranted.

B. NETWORK ARCHITECTURE

The neural network employed consisted of a three layer network utilizing the normalized cumulative backpropagation algorithm. An illustration of this preliminary network is provided in Figure 10.

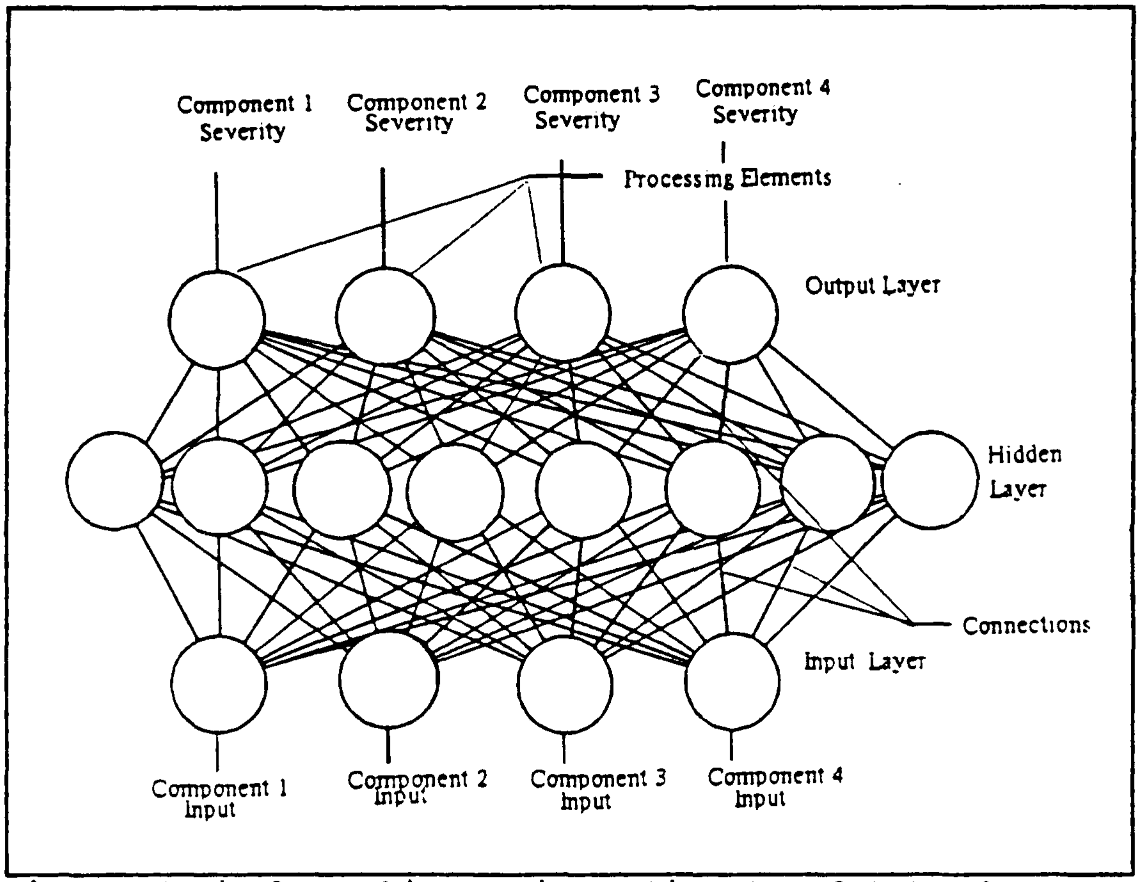

Figure 10 Simple Machinery Diagnostics Neural Network

The normalized cumulative backpropagation algorithm was selected because of its tendency to smooth out oscillations in

weight changes by adjusting weights once each epoch of vector presentations, thereby tending to minimize the global error rather than the local error associated with a single vector. While a standard backpropagation network was tried, learning became unacceptably slow with the weights and errors fluctuating wildly with little net improvement in RMS error.

All processing elements in the hidden and output layers utilized the hyperbolic tangent transfer function; the input processing elements were not influenced by a learning rule and employed purely linear transfer functions. All processing elements were connected to a weighted bias whose excitation was continuously 1.0 but whose weights could be adjusted. While the sigmoid transfer function may be currently more popular for backpropagation, the ability of the hyperbolic tangent to provide negatively signed outputs seemed advantageous for use in follow-on networks. As research continued, it was found that networks utilizing negatively signed input and output vectors had difficulty in converging satisfactorily. Consequently, this feature was ultimately not capitalized on. The layer architecture with the input processing elements not directly participating in learning and the employment of the bias element are standard features of the backpropagation algorithm[Refs.8 and 18].

The optimum number of processing elements to be used in the hidden layer was difficult to determine precisely. To obtain a better understanding of this parameter, it was

decided to verify some of the work accomplished in chemical process diagnostics by Venkatsubramanian and Chan[Ref.2] on this parameter but using networks designed for mechanical diagnostics.

C. EXPERIMENTAL PROCEDURE

Initially a training set was established by building input vectors reflecting dB differences from an established baseline at the characteristic frequencies for four ficticious machinery components. These input vectors provided generally constisted of three inputs within the dB region correlating with a severity response of zero and one corresponding to a higher severity response. Additionally sample vectors having no faults and a few vectors reflecting multiple faults were included.

This training set consisted of 48 vectors. This number of vectors was based on practical experience that it was best to use a minimum of between three to five vectors per processing element when conducting training[Ref.23]. An example of these training sets as well as a test set and a network response are included in Appendix A.

The number of processing elements in the networks investigated was based on the conventional wisdom that recommends that the hidden layer consist of between one and two times the number of processing elements in the input layer. Networks containing four,five,six, and eight processing

elements in their hidden layer were trained and tested utilizing this training set. However, since it was reported by Marko et al[Ref.5] that success was obtained using fewer hidden elements, training was attempted using three and twelve hidden elements as well.

During the training process, the number of training iterations required to reach certain discrete RMS errors were noted. While RMS error is useful in determining how close actual network response compare to desired response, it is based on the samples actually used in training. It tells nothing about what level of success can be expected when presented with new data with which it will be required to make a diagnosis. To provide an indication of this, test sets containing input vectors not previously presented to the network during training were used. Two test sets were used, one containing 15 vectors, and the other containing 16 vectors. These vectors included a number of examples near the borders of each defined severity region and a few multiple fault examples.

The "grading" of the test outputs was somewhat arbitrary. While overall RMS error experienced in the test set may have been useful, there may have been a clear separation of fault levels even though the error calculated exceeded the RMS error to which the network had been trained. Accordingly, a test grading criterion of "go" or "no go" was employed wherein an arbitrary 0.15 threshold level was established about each

desired output severity level. If the actual output vector was within the threshold, at all nodes, the network had responded correctly and received "full credit". If the actual vector output exceeded the threshold but never crossed into a region established by actual output of the network corresponding to another severity level, it was considered marginally correct and received "half credit". This reflects the fact that while it may have exceeded the threshold, no misdiagnosis had really occurred. If any other result occurred, the network received "no credit" for that particular test vector presentation.

D. EXPERIMENTAL RESULTS

A summary of the results is provided in Figures 11 and 13. Initial learning was most rapid for the six hidden element network, which reached an RMS error rate of 0.15 in 1350 vector presentations. The four, five, and eight hidden element networks took 88%, 71%, and 136% more iterations respectively to arrive at the same level of convergence. However, the six hidden element network proved slowest to improve this level of convergence to 10% RMS error, requiring 72,500 iterations compared to 33%, 20% and 41% of that number for the other networks. The three and 12 hidden element networks were used to explore the stability of the network during the early stages of learning and were not run to particular convergence levels. Consequently they are not included in Figures 11 and 13. Nevertheless it can be reported that the three hidden

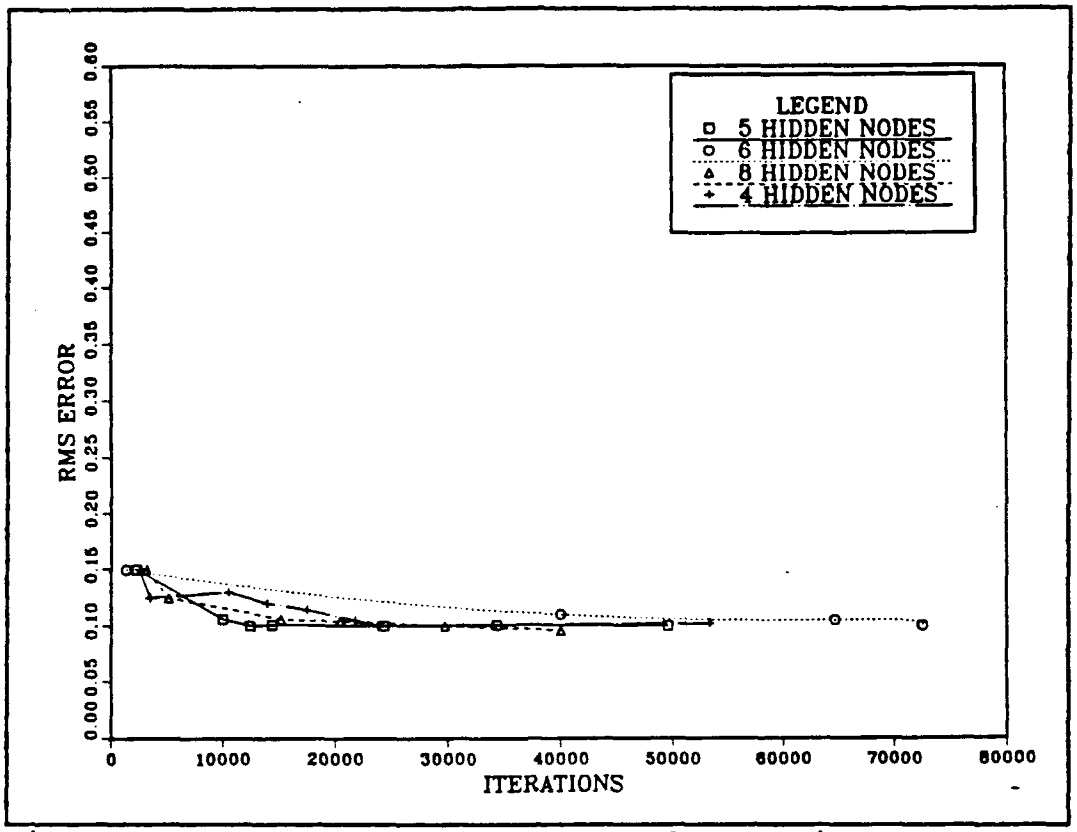

Figure 11 RMS Error Versus Number of Iterations

element network required in excess of 15,000 iterations to reach a RMS error level of 0.15. The 12 hidden element network required in excess of 36,000 iterations to reach an RMS error level of 0.25.

Observation of each network's response during the early stages of training is also noteworthy. The low hidden element networks tend to learn more rapidly at first but reach a plateau in error rate, whereafter learning is slow. At high numbers of hidden elements, the learning is characterized by a degree of instability, where RMS error levels fluctuate considerably and large errors are prone to occur. In these

networks, learning is also extremely slow from the outset, presumably due to processing elements in the hidden layer competing with one another for a limited number of features in the decision space. A sketch of the RMS errors from the start of learning to about 2000 iterations is provided in Figure 12.

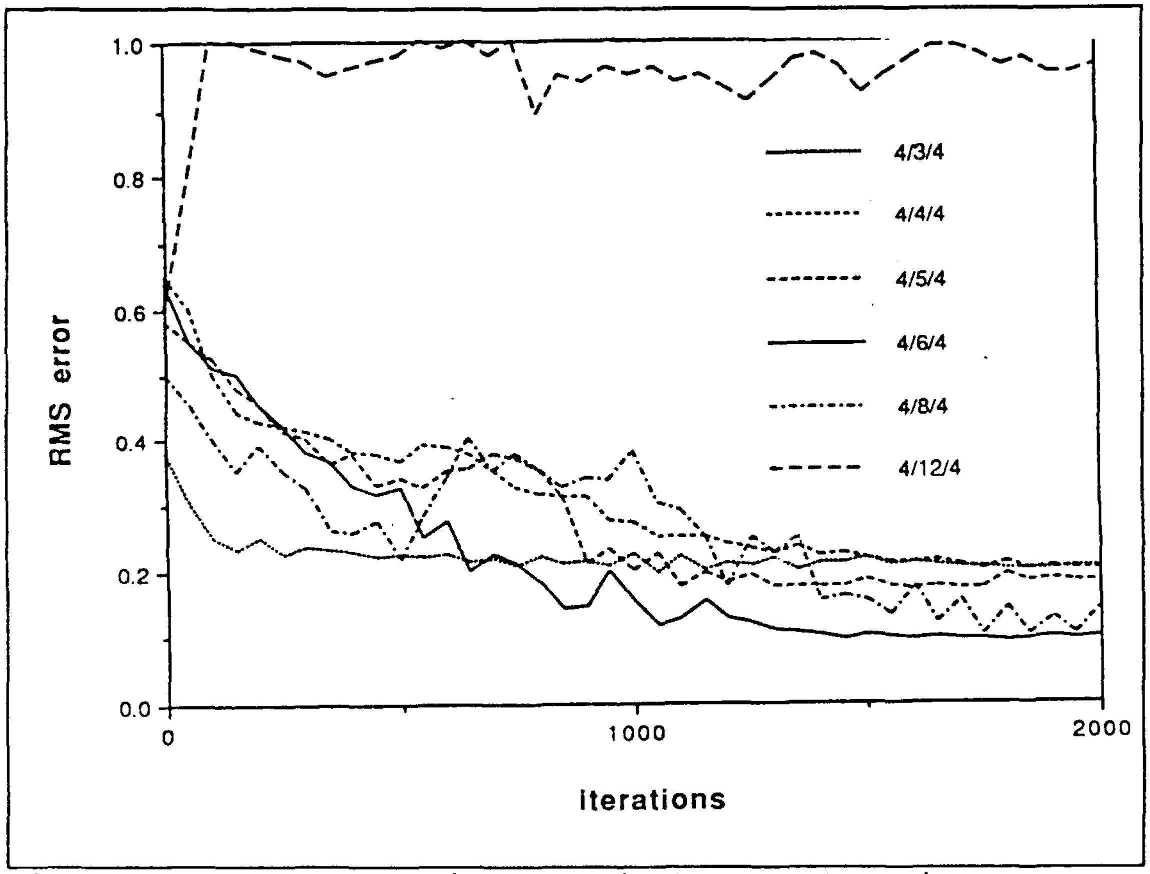

Figure 12 RMS Error During the First 2000 Iterations

Test results were similar but not identical to training RMS results. The least successful network at the same RMS level was the four hidden element network with a 71% success rate. At RMS error levels of 0.15, 83.9% successful responses were obtained by the five and eight hidden element networks.

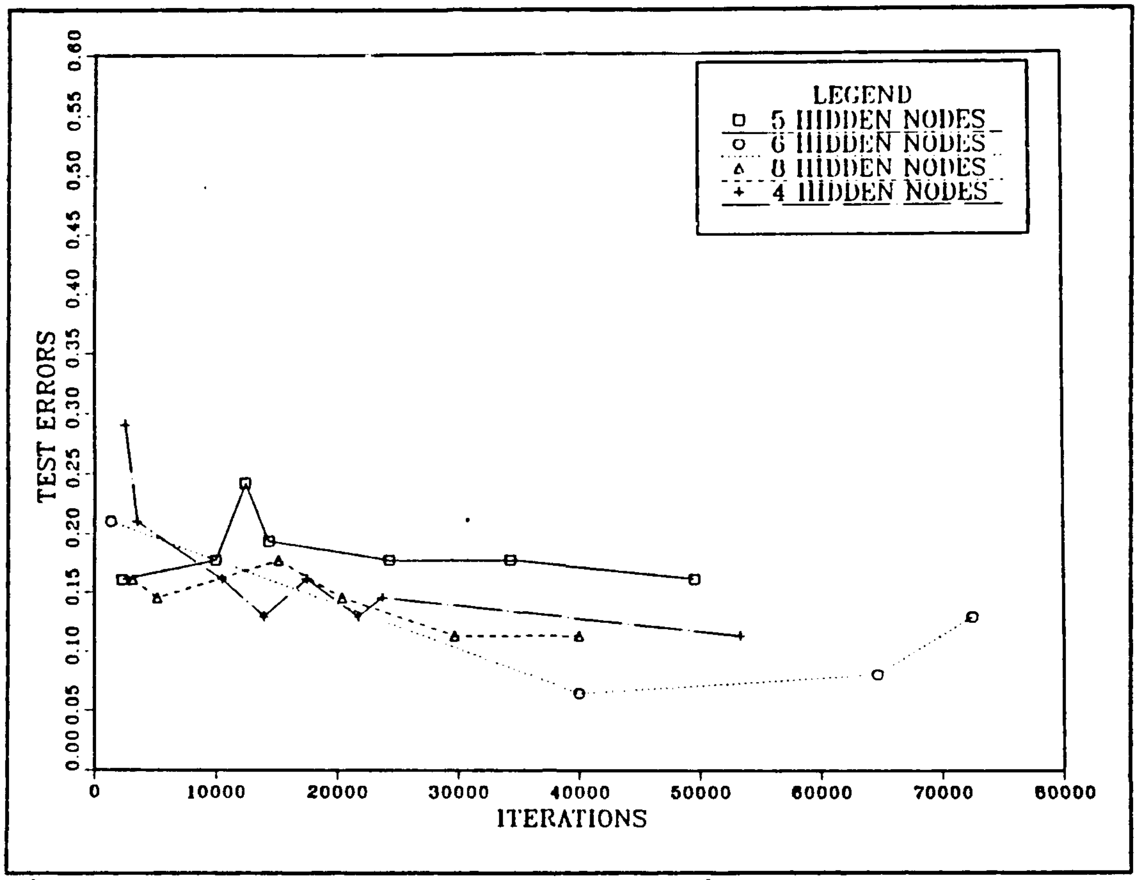

Figure 13 Test Response versus Iterations

At an RMS level of 0.10, the eight and four hidden element network improved to 87.1% while the five hidden element network remained the same. Overall, success rates improved little after an RMS error level of 0.15 had been reached; however, the bandwidth of the test responses constricted about the desired severity levels considerably, making it much easier to determine the severity level as training continued.

A major source of the errors that did occur involved the test vectors that explored the boundaries between severity levels. This is not terribly surprising as neural networks are by nature analog systems which are not particularly adept at

precise numerical calculations[Ref.19]. This is also the same region where a biological "expert" would have the greatest difficulty. In the case of the six hidden element network, the number of test errors actually increased following extensive training. This would appear to be an example of overtraining, where the pattern features of the training set become so closely mapped, that generalities associated with the actual decision space represented by the training set are missed.

As mentioned previously, several multiple fault cases were presented to the network during the testing phase. Although a few multiple faults were included in the training set as well, it is highly encouraging to observe that the networks all responded well to these multiple faults. Additionally, during one of the training phases, it was discovered that one of the input vectors had an erroneous desired output listed. The training file was corrected and learning was allowed to continue. After a number of iterations, the network in question performed as well on the previously faulty vector as on any other. This demonstrates that backpropagation networks have the ability to update themselves with new data without having to start afresh. On the other hand it also demonstrates the network's ability to forget old data if it is removed from the training set. Tables of a sample of the test sets and training sets utilized in these preliminary experiments are provided in Appendix A.

E. DISCUSSION OF RESULTS

Based on the results of the preliminary experiments delineated above, it would appear that the optimum number of hidden nodes within a certain range depends on one's priorities. If one is interested in rapid learning possibly at the expense of the level of convergence and corresponding performance on test sets or in the field, use of a minimal number of hidden elements commensurate with getting the convergence level required would be in order. In this case either four or six hidden elements would suffice. If one is more interested in accuracy rather than speed of convergence, then a higher number of hidden nodes, such as six or eight in this case, would be in order. If excessive numbers of hidden elements are used, the network tends to become unstable, as in the case of 12 hidden elements. If too few hidden elements are used the level of convergence remains excessively high and rate of convergence becomes excessively slow. However, within the range of converging networks, it would appear that the number of hidden elements is immaterial, provided that a satisfactory level of convergence is met.

The ability of the backpropagation neural networks to train on updated data without having to start afresh as well as their ability to identify multiple faults is highly encouraging, as these are both areas where conventional expert systems have some degree of difficulty.

Nevertheless, thus far, all that has been accomplished is a mapping of a dB difference to a somewhat arbitrarily derived severity level. A few lines of FORTRAN code could do the same thing. Furtermore, the use of single frequency inputs to identify machinery faults is somewhat oversimplified. A more sophisticated diagnostics model is required to determine the feasibility of neural networks in the field of machinery condition monitoring and diagnostics. Such a model is described in the following chapters. However, the neural networks so employed have their basis in the model described here.

V. DIAGNOSTIC SYSTEM PROTOTYPE: THE PHYSICAL MODEL

This chapter describes the medium complexity rotating machinery for which the diagnostic system was designed as well as the equipment utilized to monitor it. It also describes the nature of the machinery faults imposed, the portions of the vibration medium utilized for inputs for the neural network and the basis for these inputs. The procedure by which the experimental data was obtained is described and finally, the data obtained from the physical model is presented and analyzed.

To determine whether neural networks could be utilized in a machinery condition monitoring and diagnostics application, it was decided to develop a neural network diagnostic system for an uncomplicated piece of machinery that could be easily supported in a laboratory environment. This physical model would have to possess components that could be damaged with minimal expense in order to create the fault conditions for diagnosis.

A. MODEL DESCRIPTION

The medium complexity gear model utilized for these experiments was based on the machinery utilized in Robinson's[Ref.37] experiments on statistically based

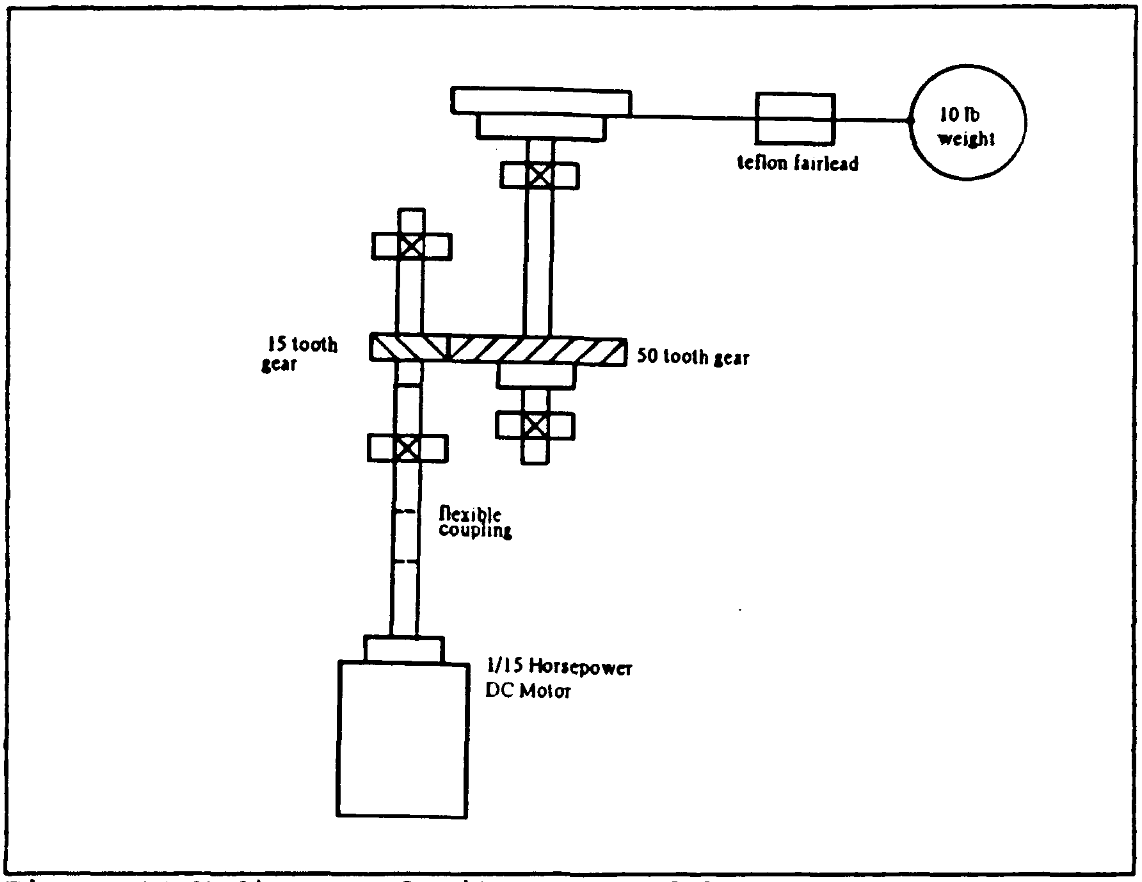

Figure 14 Medium Complexity Gear Model

vibration data. A schematic of this machinery is presented in Figure 14. It consisted of a single reduction gear train consisting of a 15 tooth drive gear (Gear 1) and a 50 tooth driven gear (Gear 2). The gears were both Martin 20 diametral pitch 3/8 inch face hubbed spur gears with a 14.5 degree pressure angle. Each was attached to a 3/8 inch diameter shaft by means of a set screw recessed in the hub which allowed for easy removal.

The shafts were each supported by two Fafnir 3/8 inch bore radial ball bearings. These bearings were mounted in aluminum block housings which were in turn bolted and glued onto a 1.0

inch thick plexiglass slab which rested on a heavy cast iron base. A vibration absorbing sheet was placed between the plexiglass and cast iron base to minimize the influence of extraneous vibrations on the system.

The drive shaft was connected to a 1/15 horsepower 0.75 Amp 115 volt variable speed DC motor by means of a rubber flexible coupling made from a piece of automotive fuel hose. The fuel hose coupling had the advantage over other flexible couplings in that it was inexpensive, easily replaced, and allowed for greater vibrational isolation between the motor and the gear train. This had the effect of improving the isolation of the gear train from vibrational influences of the motor while permitting small misalignments between the two components.

A frictional load was imposed on the drive train by means of a 3.0 inch pulley wheel which was allowed to work against a rawhide thong onto which was hung a 10 pound weight. The uniformity of the applied load was further enhanced by using a teflon fairlead to hang the weight over the side of the base, thereby reducing variable frictional effects on the rawhide thong.

Motor speed was made adjustable by means of a Bodine Electric Company combination rectifier and variable potentiometer speed controller. This simple feed-forward speed controller was manually adjusted to the desired speed of operation by metering shaft RPM's with a Power Instruments

Model 1720 RPM indicating optical proximeter. In these experiments, shaft speed was maintained at as near to 30 Hz as possible.

B. VIBRATION MONITORING EQUIPMENT

The principle components of the vibration monitoring suite used in this experiments were a PCB Model 303A03 accelerometer, a PCB Model 480D06 accelerometer power supply, a Hewlett Packard Model 3562A Dynamic Signal Analyzer, a Iwatsu Model SS5702 20 MHz Dual Channel Oscilloscope, a Gould Type 1421 20 MHz Digital Storage Dual Channel Oscilloscope, and a Hewlett Packard Model 7035B X-Y recorder. A schematic of their arrangement is provided in Figure 15.

1. PCB Model 303A03 Accelerometer

The PCB Model 303A03 Accelerometer is a medium range high frequency miniature accelerometer, based on a piezoelectric quartz transducer sensing element. This accelerometer possesses the following parameters:

- Sensitivity: 10 mV/g
- Resonant Frequency: 70 kHz
- Range: ± 500 g
- Resolution: 0.02 g
- Size: 0.28 X 0.4 in
- Weight: 2.0 gm

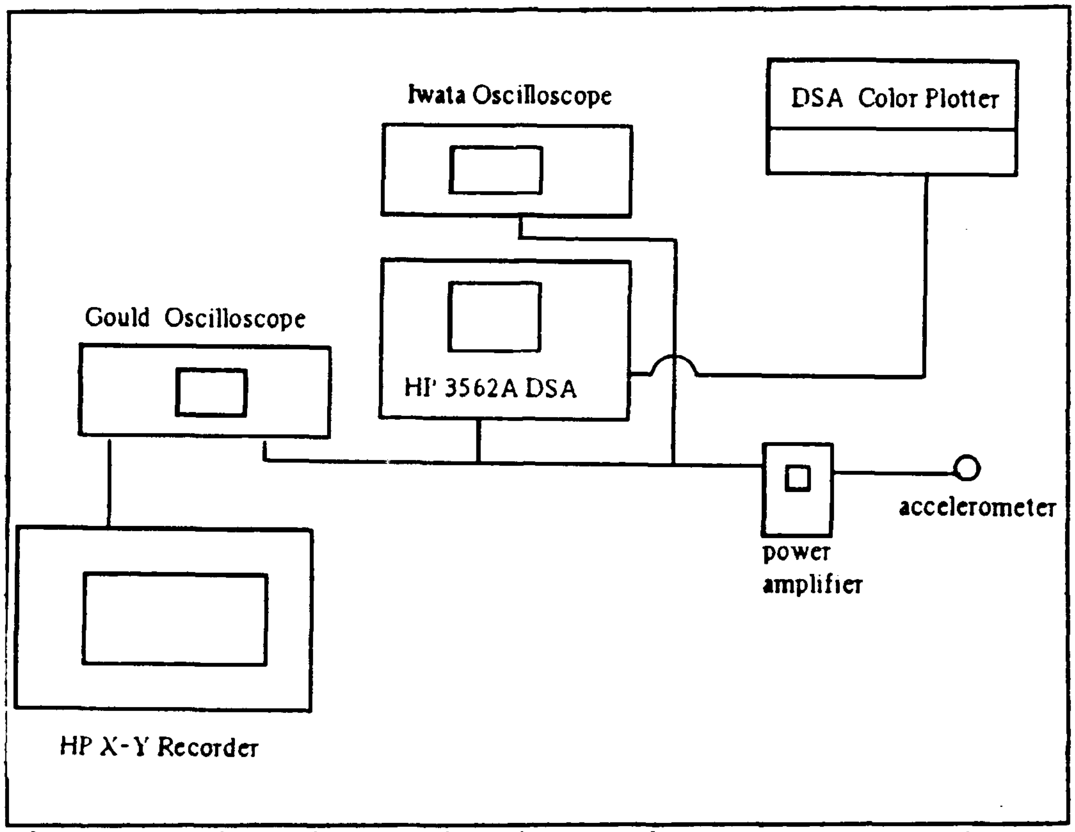

Figure 15 Vibration Monitoring Equipment Arrangement

The accelerometer was mounted in a radial position directly above the bearing supporting the shaft driven by the 50-tooth gear closest to the gear itself. It was affixed to a permanently attached mount by means of mounting wax and thereby was not itself permanently affixed.

The accelerometer output voltage was amplified by a PCB Model 480D06 power supply which provided a DC power source with which to amplify the signal. During the entire experiment this power supply was set up to amplify the accelerometer output by a factor of 10.0.

2. Hewlett Packard 3562A DSA

The heart of the vibration monitoring system was the HP 3562A Dynamic Signal Analyzer (DSA). This is a dual channel FFT analyzer capable of measuring the complete spectrum of vibration parameters, including time domain and statistical parameters as well as the more traditional linear and power frequency spectra. It is also capable of a large number of mathematics functions, including the performance of the logarithmic functions and inverse Fourier transforms required for Cepstral analysis. In these experiments, the DSA was primarily used to measure the linear frequency spectrum from 0.0 to 1500 Hz and the Cepstrum over a similar range. The baseline parameters utilized during these experiments are provided in Figure 16.

3. Peripheral Equipment

A number of time domain monitoring and plotting devices were used alongside the 3562A DSA. Because the DSA is somewhat restricted in the length of time signal that can be measured at a given time due to time record length constraints inherent to the FFT, a Gould 1421 recording oscilloscope was utilized in conjunction with a HP 7035B X-Y plotter to record time signals of interest whose features warranted a time length other than that of the time record.Additionally, an Iwatsu SS5702 Oscilloscope was substituted for the Gould to provide an additional means to observe the time signal while

Linear Resolution

MEASURE:	CHAN 1		CHAN 2	
	Freq Resp		Freq Resp	
WINDOW:	CHAN 1		CHAN 2	
	Hanning		Hanning	
AVERAGE:	TYPE	# AVGS	OVERLAP	TIME AVG
	Stable	15	0%	Off
FREQ:	CENTER		SPAN	BW
	156.25 Hz		312 Hz	586mHz
	REC LGTH	Δt		
	2.56 S	1.25mS		
TRIGGER:	TYPE	LEVEL	SLOPE	PREVIEW
	Freerun	0.0 Vpk	Pos	Off
INPUT:	RANGE	ENG UNITS	COUPLING	DELAY
CH 1	AutoRng↑	1.0 V/EU	DC (Flt)	0.0 S
CH 2	AutoRng↑	1.0 V/EU	DC (Flt)	0.0 S
SOURCE:	TYPE		LEVEL	OFFSET
	Rndm Noise		0.0 Vpk	0.0 Vpk

Figure 16 Experimental Operating Parameters for the HP3562A DSA

the HP 3562A DSA was otherwise occupied. Additional accessories to the 3562A DSA which proved invaluable during the data acquisition and storage phases of the experiment were the HP color pen recorder and HP 9122 hard disc drive.

C. DETERMINATION OF MONITORED PARAMETERS

By far the most critical decisions in this study involved a determination of the vibration parameters to use as inputs to the neural networks. In order for the network to perform its task adequately, two things must occur. First, the dimension of the input vector and the corresponding number of

input PE's must be sufficient to thoroughly describe the decision space which the network is tasked with categorizing, whether this be a range of signal pattern or machinery diagnostic faults. Secondly, especially in the case of performance based learning algorithms such as backpropagation, the training data must be sufficiently varied to reflect the range of decisions expected of the network and must be of reasonably good quality. The neural network may be categorically tolerant of noisy data, but it is still subject to the adage, "garbage in, garbage out." Additionally, the computational load imposed by the neural network during the training phase is a function of the number of processing elements involved, and thus indirectly is a function of the number of inputs. Therefore it is desirable to keep the dimension of the input vectors to the minimum necessary to describe the decision space.

In Chapter III it was stated that in this research the machinery faults of particular interest were those associated with the gears, bearings, and shaft misalignments. This is also the limit of the rotating components available in the uncomplicated machinery under investigation. The choices of inputs therefore were restricted to parameters associated with these components.

The question of which medium to employ as the principal source of inputs was critical. Robinson[Ref.37] found that statistical measurements of the time domain were superior to

those of the frequency domain for the detection of machinery faults, especially gear faults. This is corroborated by the work of Matthew and Alfredson[Refs.26 and 38] state that time averaged signals and matched filtered spectral signals should be capable of detecting gear anomalies long before conventional spectral analysis. However, the main thrust of this work concerns isolating the location of the fault, which is much more directly accomplished in the frequency domain, unless a long series of different filtered time signals are used. As this was once the method of measuring the frequency domain before the advent of the Fast Fourier Transform, this really is just another form of spectral analysis. Additionally, while the HP 3562A DSA is capable of statistical time domain analysis, it is better suited to analysis of the frequency spectrum. Further, to measure statistical parameters in the time domain, the DSA requires the use of an accurate RPM indicator to provide a trigger signal. Although the proximeter in use to measure shaft speed was sufficiently accurate to provide a trigger signal, it tended to become erratic when having difficulty in establishing an optical reference. As a result, time domain statistical parameters were not employed as inputs to the diagnostic system. However, during the data acquisition stage, some time domain signals were recorded for reference. Consequently, the frequency spectrum was used as the primary source of diagnostic information.

Determining the frequency inputs for the gears was fairly straightforward. Randall[Ref.38] recommended that monitoring in the vicinity of the first three harmonics of the gear mesh frequency would provide for earliest detection of uniform wear gear faults. With the physical model operating at 30 Hz, using equation (21), the gear mesh frequency was calculated to be 450 Hz.

It is also well known that damage to the gears is most often characterized by the growth of the sidebands associated with the rotating frequencies of the gears within the mesh. There are many suggested methods of representing this. One such method involved observing the magnitude of the spectrum one shaft rotation frequency up and down from the gear mesh frequency. This took into account the observation that the first sideband seemed most sensitive to gear damage. Another proposed method involved integrating the frequency spectrum and taking the limits of integration from one or two sidebands on either side of the gear mesh frequency. This took into account the idea that the severity of the fault was proportional to the energy level of the frequency response of the system. A final possibility is to simply take the average of the first three sidebands associated with each gear on either side of the gear mesh frequency. This has the advantage of being easier to calculate than an integral and yet is essentially a normalized integral. Further, it takes into account the existence of more than the first sideband and

tends to add stability with respect to successive measurements. As the input into the neural network is based on the dB difference from a baseline and is relativistic in nature, the averaging does not detract from its utility and offers an excellent compromise between the other two options.

Randall[Ref.34] reported good results in the use of cepstral analysis in gear diagnostics and presented several practical points in its implementation. As the effect of the cepstrum is to compress whole families of harmonic frequencies into a single quefrency and perhaps one or two rahmonics, it seems an ideal parameter by which to identify sideband growth. Thus the quefrencies associated with the 9.0 and 30 Hz sidebands were employed as alternative inputs to the averaged sidebands obtained from the frequency domain.

Bearing parameters were somewhat more difficult to come by. While impact frequencies for the inner and outer race as well as the balls themselves are easily derived, they invariably occur at low frequencies, where they are obscured by the higher energy impacts associated with the gears as well as extraneous noise. As a result, it is recommended that one look to high frequency harmonics for this information. Regrettably, in preliminary sweeps of the frequency spectrum up to 3000 Hz, no high frequency signals associated with the bearings were detected. This is probably a result of the small size and light loading of the particular bearings involved. Nevertheless, some weak signals were noted at the first and

second harmonics of the inner and outer race impact frequencies. As a result, these frequencies associated with the 30 Hz shaft bearings, as well as the ball impact frequency were monitored in the hopes that something might become discernible when a bearing casualty was imposed.

As in the case with gears, bearing impacts are readily discernible on analysis of the cepstrum. Van Dyke[Ref.35] reported excellent results with cepstral analysis on the detection of bearing faults in maritime propulsion plant and auxiliary machinery aboard U.S. Navy aircraft carriers. Consequently the quefrencies associated with the 9.0 Hz shaft bearings were also monitored.

Collacott and several others[Refs.27 and 31] indicate that the bulk of shaft imbalances and misalignments are detectable at between 0.5 and 2.0 times the shaft rotative frequency. Consequently, the first two harmonics of each shaft were monitored.

In summary, the following frequencies and quefrencies were monitored.

- The gear mesh frequency and the next two harmonics; 450, 900, and 1350 Hz.
- The average of the first three of the 9.0 and 30 Hz upper and lower sidebands surrounding the gear mesh frequency and its harmonics.
- The cepstral quefrencies associated with the 9.0 and 30 Hz sidebands; that is, 33.3 and 111 ms.
- The average of the cepstral rahmonics associated with the sidebands where available; that is, 33.3 ms and its next two rahmonics.

- The first two harmonics of the 30 Hz shaft bearing inner race defect and outer race defect frequencies; that is, 118, 236, 92, and 184 Hz.
- The 9.0 Hz shaft bearing ball defect frequency;that is, 103 Hz.
- The bearing related quefrencies, 8.5, 9.7, and 10.9 ms.
- The average of the first three rahmonics of the 10.9 ms quefrency.
- The shaft rotative frequencies and their next harmonics, 9.0, 18, 30, and 60 Hz.

Several additional frequencies were recorded as their prominence became apparent. However, as these frequencies were not recorded in all of the experiments, they were not utilized as inputs to the neural networks that follow.

D. DATA ACQUISITION PROCEDURE

The physical model was utilized to extract the frequency spectral and cepstral data delineated in the previous section. The first tests were conducted over the period of several days with all mechanical components in their normal operating condition in order to establish a baseline. The machinery components were then systematically subjected to damage with one new perturbation per test. In each test, the following general procedure was adhered to.

Prior to any data extraction, any new machinery components to be employed in the test were worn in over several hours at the operating speed of 30 Hz. This was particularly necessary for the gears whose associated parameters would vary from

reading to reading until they were worn in and all blacking had been removed from the gear tooth contact surfaces.

In addition to this wear-in time, which was only imposed on tests involving new components, all tests were subjected to a mandatory 45 minute stabilization period during which no parameters were recorded. This was determined to be a sufficient time period for the machinery to reach a state where the parameters monitored became statistically stable and the readings largely became repeatable to within 3.0 dB.

Following the stabilization period, the recording of parameters began. Although the Gould digital storage recorder was not available throughout the experiment, it was utilized extensively when available to record time domain signatures in conjunction with the HP X-Y plotter. This was used to record any portion of the time signal that may have been of interest.

Following recording of the time signal, a series of narrow band linear spectrum plots were obtained using the DSA and its color pen recorder. All parameters recorded from the DSA utilized a stabilized mean with 15 averages. The narrow band linear spectrum plots covered the pertinent sections of a broad band region from 0 to 1535 Hz. Specifically, recordings were taken with a frequency band of 312 Hz with starting frequencies of 0, 300, 750, and 1200 Hz. Following this a broad band power spectrum was obtained with a frequency span of 1535 Hz. The log of this plot was then taken followed by an Inverse Fourier Transform, resulting in a broad band cepstrum.

This was performed automatically using the Cepstrum function of the DSA.

During the first set of readings in a given test, plots of all frequency spans and the cepstrum were recorded. Subsequent readings were not accompanied by recorded plots; only the parameters of interest were recorded. A total of between six and eight of these sets of readings would be taken in a given test to ensure a statistically stable data base and to establish a larger number of sample vectors with which to test and train the neural networks. As a result of the procedures delineated above, each test took approximately four hours to accomplish.

Following the recording of the entire test set, means and sample population standard deviations were computed for each parameter obtained. The purpose of this was twofold. First the statistical parameters allowed a judgement to be made about the stability of the data and consequently its repeatability. It was also hypothesized that the variance in the standard deviation of the readings could be indicative of the severity of the impacts at that frequency and thus could prove to be a useful diagnostic tool. Secondly, observation of the mean of each of the parameters enabled comparisons between tests to be made at a glance, thereby providing an indication of how well the parameters could be expected to represent the diagnostic decision space for the model.

E. PRESENTATION OF EXTRACTED DATA

A total of twenty two test sets were conducted using the simple gear train model. Of these three sets involved entirely undamaged machinery and were used to establish the baseline and provide data for "normal" equipment readings. Nine tests were conducted with various levels of damage imposed on the 15 tooth pinion, hereafter identified as "Gear 1". Four tests were conducted with various levels of damage imposed on the 50 tooth gear, hereafter identified as "Gear 2". One test was conducted with damage imposed on both gears. Two tests were conducted involving bearing damage and three tests were conducted involving shaft imbalance and misalignment. These tests are summarized in Table II.

1. Tests Involving Undamaged Equipment

A number of tests were conducted to establish a baseline but ultimately only three of these test sets were utilized in the neural networks. These tests featured a rather wide range of amplitudes in spite of the efforts to allow the system to stabilize. In fact, the variation of normal readings would appear to exceed that of damaged machinery by a significant margin.

Figure 17 illustrates the time signal for an undamaged machine.Figures 18 through 21 illustrate a sample set of 312 Hz span linear spectra for normal machinery. Figure 22 illustrates the broad band cepstrum for the undamaged

Table II. Summary of Tests Performed on Physical Model

Test	Shaft 1	Shaft 2	Brng. O.R.	Brng. I.R.	Ball Def.	Gear 1	Gear 2
Nor 1	-	-	-	-	-	-	-
Nor 2	-	-	-	-	-	-	-
Nor 3	-	-	-	-	-	-	-
G1-1	-	-	-	-	-	Mod	-
G1-2	-	-	-	-	-	High	-
G1-3	-	-	-	-	-	Mod	-
G1-4	-	-	-	-	-	Low	-
G1-5	-	-	-	-	-	Low	-
G1-6	-	-	-	-	-	Mod	-
G1-7	-	-	-	-	-	Mod	-
G1-8	-	-	-	-	-	High	-
G1-9	-	-	-	-	-	High	-
G1-10	-	-	-	-	-	High	Mod
G2-1	-	-	-	-	-	-	Mod
G2-2	-	-	-	-	-	-	High
G2-3	-	-	-	-	-	-	Low
G2-4	-	-	-	-	-	-	Mod
B1	-	-	Low	Low	Low	Mod	Low
B2	-	-	Low	Low	Low	-	-
S1-1	Mod	-	-	-	-	-	-
S2-1	-	High	-	-	-	Low	-
S2-2	-	High	-	-	-	-	-

Low: Low Severity; Mod: Moderate Severity; High: High Severity; G=Gear; S=Shaft; B=Bearing

machine.The frequency spectra are accompanied by the time record plots from which they were derived. In the frequency

spectra the gear mesh frequencies as well as numerous sidebands for both the 9.0 Hz and 30.0 Hz gears are readily identifiable.

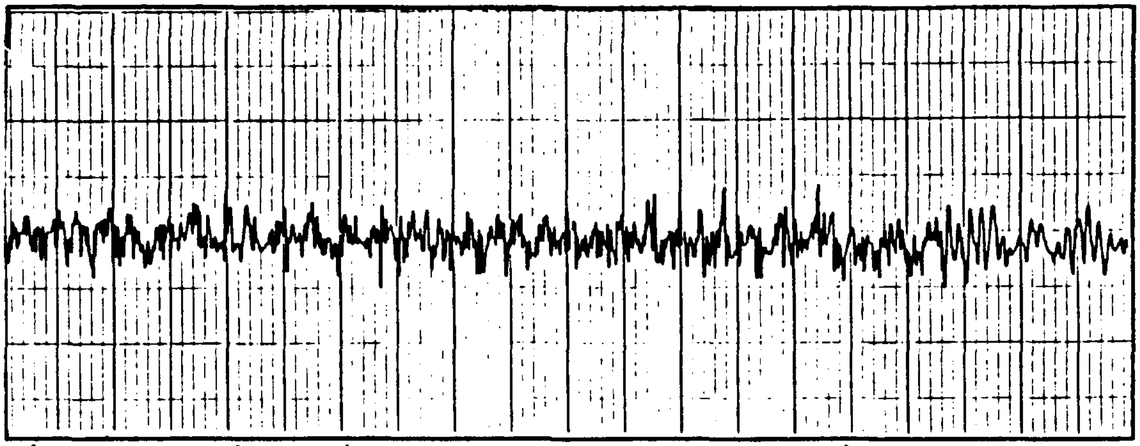

Figure 17 Time Signal for Undamaged Machine 5V/5ms per Division

Additionally, there are dominant signals at 30, 90, 180, and 270 Hz visible on the 0 to 300 Hz plot, Figure 18. The dominant signals at 90 and 180 Hz had a tendency to obscure the first two harmonics of the bearing inner race, thereby reducing its effectiveness in diagnosing bearing faults. However, as these frequencies turned out to be resonant frequencies for the system, they provide a good indication of the overall degree of excitation of the system. As a result, these particular readings were retained for the neural networks even though their utility in identifying bearing faults became increasingly doubtful as the experiments wore on.

A note concerning the appearance of the time records in Figures 5 through 8 is in order. The periodicities noted in

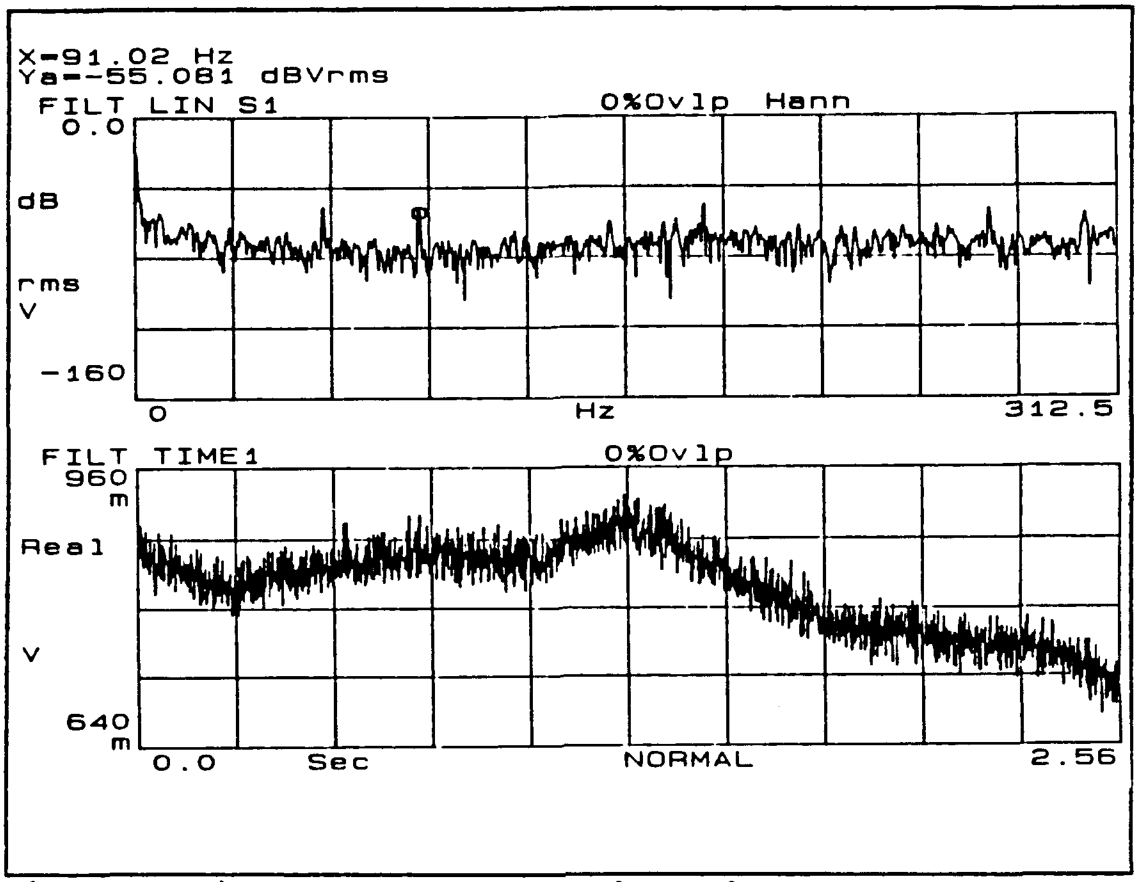

Figure 18 Linear Spectrum for Undamaged Gear 0-312 Hz

the higher frequency time records correspond to the 9 and 30 Hz sidebands. While the shaft rotative frequencies may be filtered out of the signal, these sidebands are not, resulting in the peculiar appearance of the time records.

The results of these tests are summarized in Table III. A baseline was established by obtaining the average of the first two test sets. The baseline standard deviation was based on the propagation of error formula,

$$\sigma_{total} = (\sigma_1^2+\sigma_2^2+\ldots\sigma_n^2)^{\frac{1}{2}} \tag{32}$$

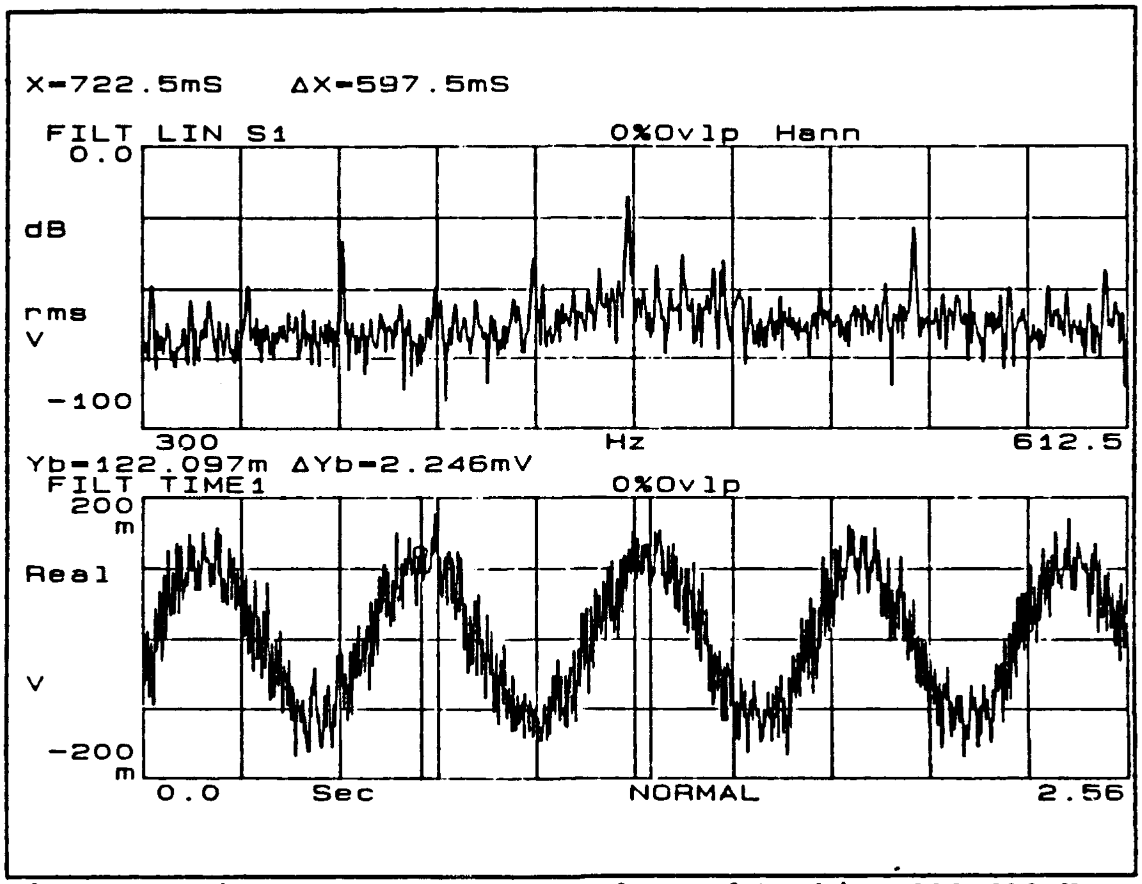

Figure 19 Linear Spectrum for Undamaged Machine 300-612 Hz

This baseline standard deviation was used as a threshold for the first severity level in the in the moderate complexity diagnostics model described in Chapter VI in the same manner as the 6.0 dB rule mentioned in Chapter IV. Severity levels for moderate and severe damage levels were generated using the largest test standard deviation involved or a value of 2.0 dB, whichever was larger. A summary of this baseline data is provided in Table IV.

Establishing a severity rating for the faults actually imposed on the various machinery components became a rather delicate task. Although establishing a severity criterion

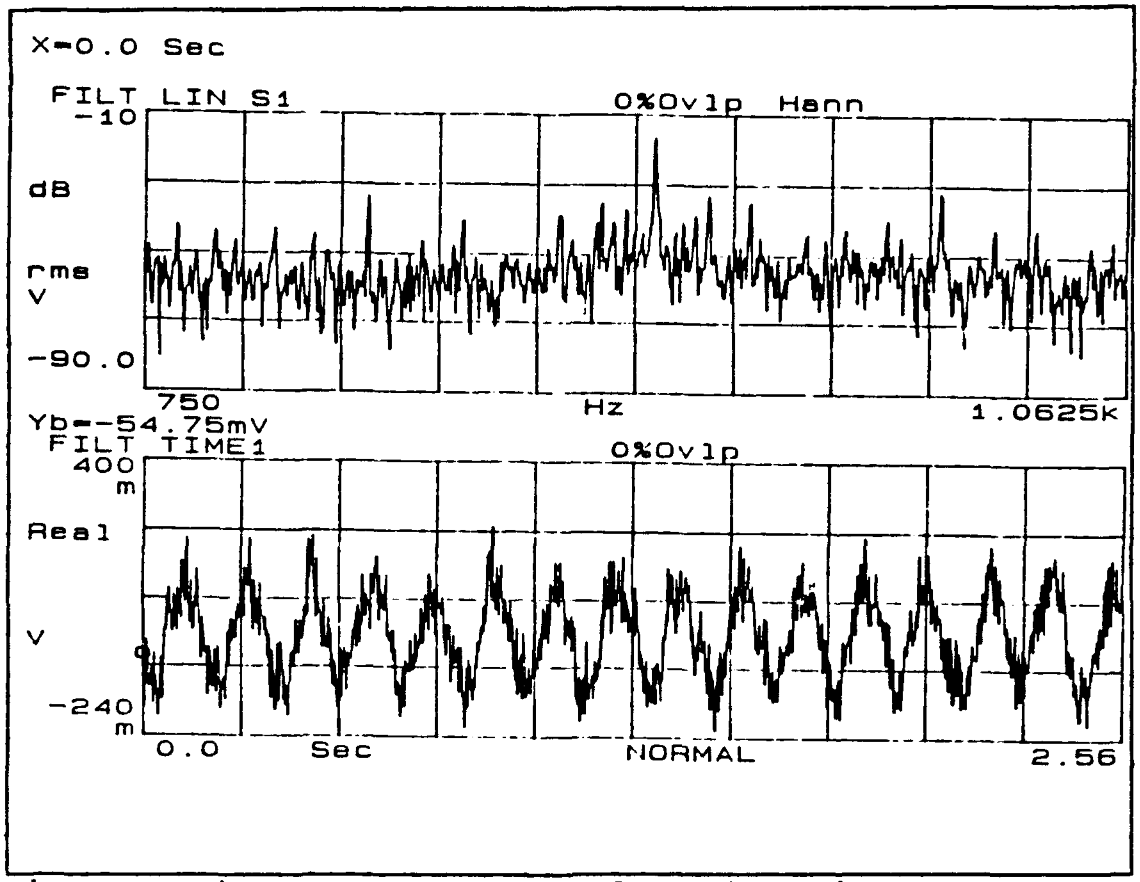

Figure 20 Linear Spectrum for Undamaged Machine' 750-1062 Hz

after the measurements were taken based on the recorded dB differences was considered, it was feared that this methodology would be analogous to fitting the data to match the theoretical model, which is not good practice . This methodology would also run counter to the purpose of a machinery diagnostics system, which is to determine the severity and location of the actual fault, and not merely its symptoms. As a result, severity levels loosely based on the extent of the physical damage were established. If, in the author's estimation, the damage was severe enough to warrant replacement at first opportunity, a severity rating of

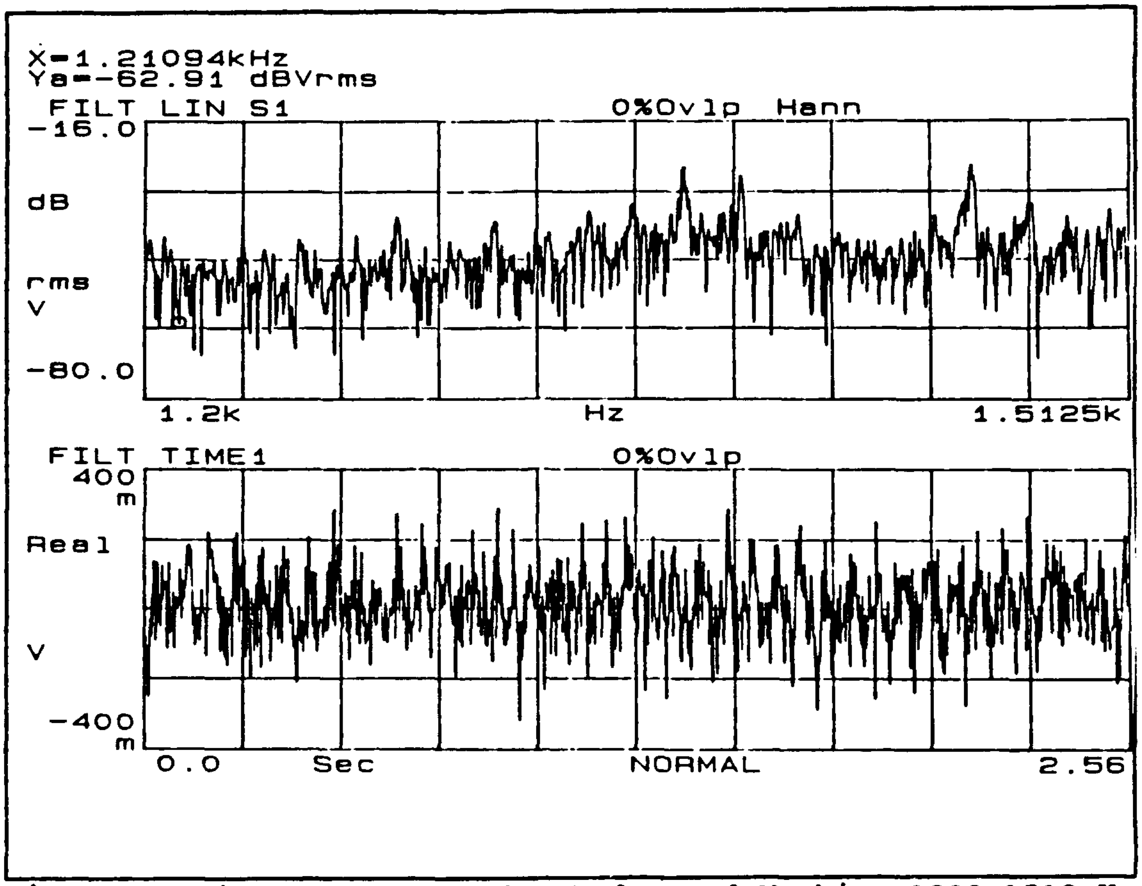

Figure 21 Linear Spectrum for Undamaged Machine 1200-1512 Hz

"severe" was determined. If the level of damage was sufficient to warrant replacement at the next scheduled maintenance period, then a severity rating of "moderate" was established. If the fault condition existed but was sufficiently light to warrant continued operation with an increased level of monitoring, a severity rating of "light" or "low" was provided. For example, if a gear tooth was completely broken off, a severe damage rating was assigned; if a gear tooth had wear inflicted such that the involute shape was just barely affected, a low severity rating was assigned. These severity levels may seem rather arbitrary but, when due consideration

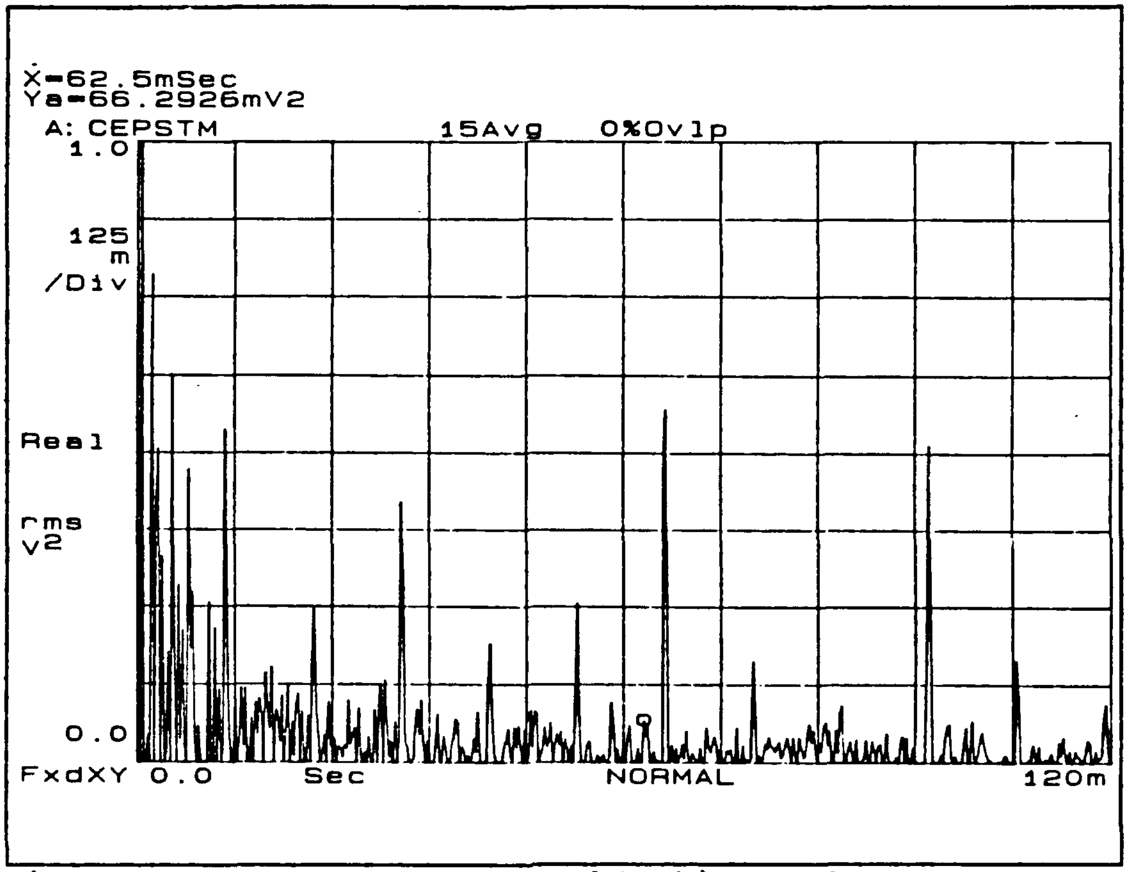

Figure 22 Cepstrum for Undamaged Machine 0-1535 Hz

of the difficulty in equating a specific degree of physical damage in a gear with a similar level of damage in a bearing or shaft, this methodology is the only plausible solution.

A quick view of Table III reveals a significant variation between the undamaged machine in Test 3 and the other two tests. This is due to the replacement and wear in of two new gears following a machinery casualty. In keeping with standard practice following a major overhaul of a machine, a new baseline was established at this point for subsequent measurements based solely on this test.

Table III. Summary of Means and Standard Deviations for Vibration Amplitudes(dB) for Undamaged Machinery

Frequency (Hz)	9.0	18	30	60	92	118	184	236	103	450	9SB
Normal No.1	-60.0 2.5	-58.5 1.2	-64.9 2.0	-63.6 2.4	-53.2 0.5	-70.5 1.4	-44.2 2.6	-63.8 4.2	-73.2 1.1	-18.8 0.5	-48.0 1.0
Normal No..2	-62.2 4.3	-64.8 2.4	-61.2 1.8	-52.7 2.5	-56.6 2.7	-70.9 3.0	-41.7 4.1	-66.6 3.1	-74.3 1.4	-14.4 1.8	-43.7 1.5
Normal No.3	-61.8 2.6	-63.9 3.8	-67.2 1.0	-48.3 1.5	-69.4 2.6	-74.0 2.3	-52.0 5.1	-66.9 2.7	-74..4 1.0	-16.3 1.2	-33.4 1.2

Frequency (Hz)	30SB	900	9SB	30SB	1350	9SB	30SB
Normal No.1	-41.4 1.9	-18.7 0.5	-39.9 1.1	-37.8 1.7	-19.0 1.2	-35.4 1.8	-33.2 1.3
Normal No.2	-38.9 1.5	-17.4 1.8	-40.7 1.1	-36.9 1.0	-19.1 3.2	-32.7 2.3	-32.4 0.8
Normal No.3	-41.0 1.2	-17.1 1.6	-32.3 1.4	-39.9 1.6	-25.4 1.9	-30.4 1.8	-34.0 1.0

SB = Sideband Average

Av = Average of first three rahmonics

Quefrency (ms)	9.7	8.5	10.9	10.9Av	33.3	33.3Av	111
Normal No.1	-7.3 0.2	-6.8 0.5	-4.2 0.5	-5.9 0.4	-4.6 0.7	-5.1 0.4	-6.6 1.0
Normal No.2	-7.6 0.6	-5.9 0.3	-3.2 0.4	-4.9 0.6	-4.1 0.5	-3.7 0.5	-6.7 0.8
Normal No.3	-7.7 0.5	-5.3 0.3	-6.0 0.6	-7.8 0.4	-7.6 0.5	-7.8 0.4	-10.9 1.4

2. Faults to the Drive Pinion

The first and most comprehensive series of tests conducted involved imposing progressively more severe damage on the 15 tooth drive pinion which was operating at the nominal speed of 30 Hz. These tests loosely followed the procedural pattern established by Robinson[Ref.37] during his work on statistical parameters in machinery diagnostics.

a. *Description of Damage*

The first test conducted involved an almost vertical filing down of the engaging face and flank of a single tooth of the drive pinion and a shallow second cut

Table IV. Baseline Decibel Levels

Frequency (Hz)	9.0	18.0	30.0	60.0	92.0	103	118	184	236
Baseline 1	-61.2 4.5	-64.6 4.5	-57.6 4.1	-52.7 2.5	-55.0 4.1	-74.4 2.0	-70.8 3.6	-41.2 6.6	-64.4 5.9
Baseline 2	-61.8 2.6	-63.9 3.8	-67.2 1.0	-48.3 1.5	-69.4 2.6	-74.4 1.3	-74.0 2.3	-52.5 5.1	-66.9 2.7

Frequency (Hz)	450	9SB	30SB	900	9SB	30SB	1350	9SB	30SB
Baseline 1	-17.4 4.1	-45.5 3.4	-40.2 3.0	-18.1 1.9	-39.6 2.9	-37.4 2.0	-19.1 3.2	-34.1 2.9	-32.8 1.5
Baseline 2	-16.3 1.2	-33.4 1.2	-41.0 1.2	-17.1 1.6	-32.3 1.4	-39.9 1.6	-25.4 1.9	-30.4 1.8	-34.0 1.0

Cepstrum (ms)	8.5	9.7	10.9	10.9Av	33.3	33.3Av	111
Baseline 1	-6.4 0.6	-7.7 1.0	-3.7 0.6	-5.4 0.7	-4.4 0.9	-4.4 0.7	-6.7 1.0
Baseline 2	-5.3 0.3	-7.7 0.5	-6.0 0.6	-7.8 0.4	-7.6 0.5	-7.8 0.4	-10.9 1.4

parallel to the top land of the gear tooth. The second test involved a deep cut parallel to the tooth base, resulting in almost complete removal of the tooth. In this case there was essentially no contact between the tooth and the driven gear. The third test involved the placement of gouges on the upper surface of two of the teeth with a depth of 1/32 inch and a width of up to 1/16 inch. These tests are identified for future reference as Gear Tests 1-1, 1-2, and 1-3, respectively. Gear Tests 1-1 and 1-3 were considered to involve "moderate" wear while Gear Test 1-2 was considered to involve "severe" wear. A schematic illustration of these damage levels is presented in Figure 23.

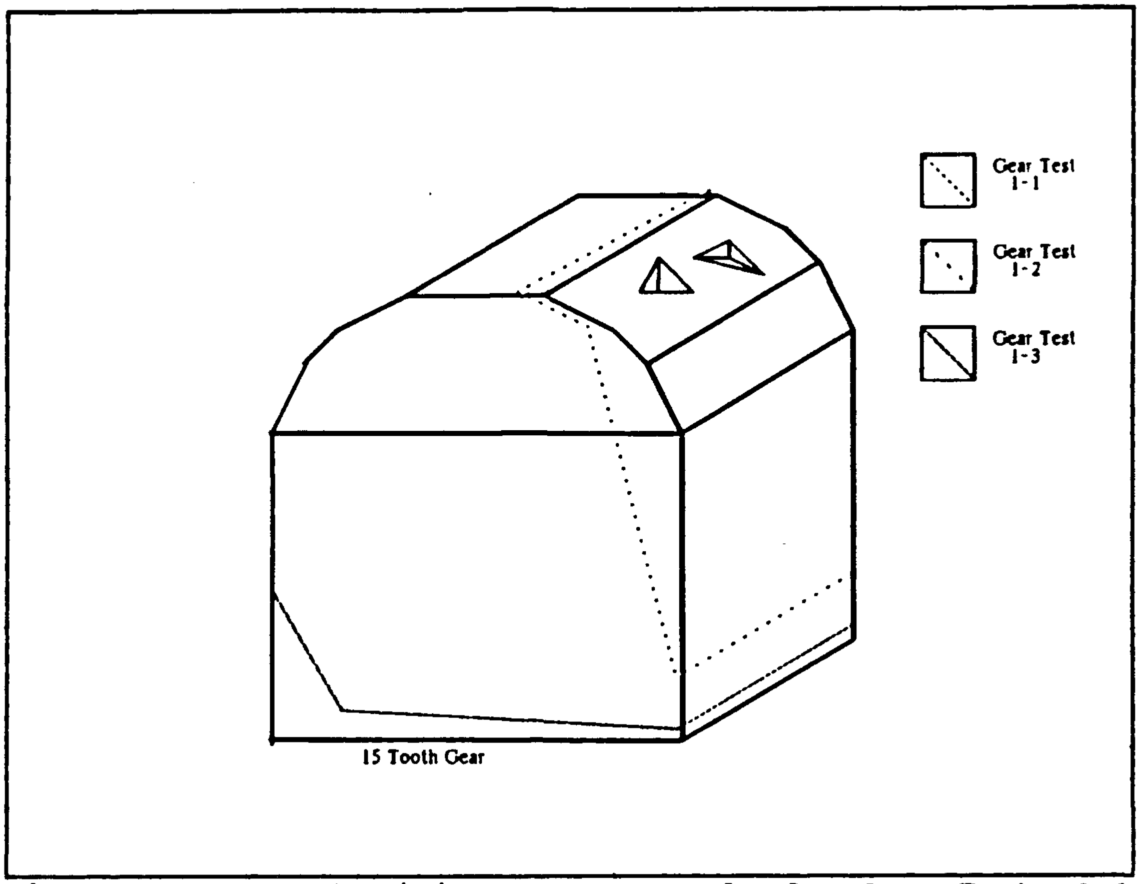

Figure 23 15 Tooth Pinion Damage Levels for Gear Tests 1-1 through 1-3

Following these tests, a more thorough set of six tests were conducted on the 15 tooth pinion. In these tests the damage was more progressive in nature. In the first test, Gear Test 1-4, a single pass was made over the engaging face of the affected tooth with a coarse machine file. Even after the 45 minute stabilization time there was a significant change in the vibration signature. However, there was no observable increase in the audible noise level from that encountered in the baseline tests. When the test

was completed, the file marks from the two passes had been removed by the wearing in phenomenon common to gears.

In the following test, Gear Test 1-5, additional material was removed from the face of the engaging face of the gear tooth but not yet biting into the top land. In this case, there was an additional clicking noise audible. Again, following the four hour testing period, the file marks had been removed except in an area on a corner where the filing had been uneven. Gear Tests 1-4 and 1-5 were evaluated as having "slight" damage.

Gear Tests 1-6 and 1-7 involved "moderate" damage to the tooth. In Gear Test 1-6, the contact surface of the engaged face was filed down until the involute shape of the tooth was clearly affected but not to the point that the top land was affected. When this test was conducted a significantly stronger clicking noise was heard. Again, no etch marks were observed following the test. Gear Test 1-7 involved deepening the region removed in the previous test until the top land was clearly affected. No additional noise during the test was noted.

Gear Tests 1-8 and 1-9 involved "severe" damage to the tooth. In Test 1-8, the removed region was deepened so that almost 1/3 of the tooth was missing. In Test 1-9 approximately 1/2 of the tooth was removed. The overall noise level during these tests increased somewhat over that encountered during the previous two tests but there was no

discernible difference between these two tests. Figure 24 depicts the damage levels for Gear Tests 1-4 through 1-9.

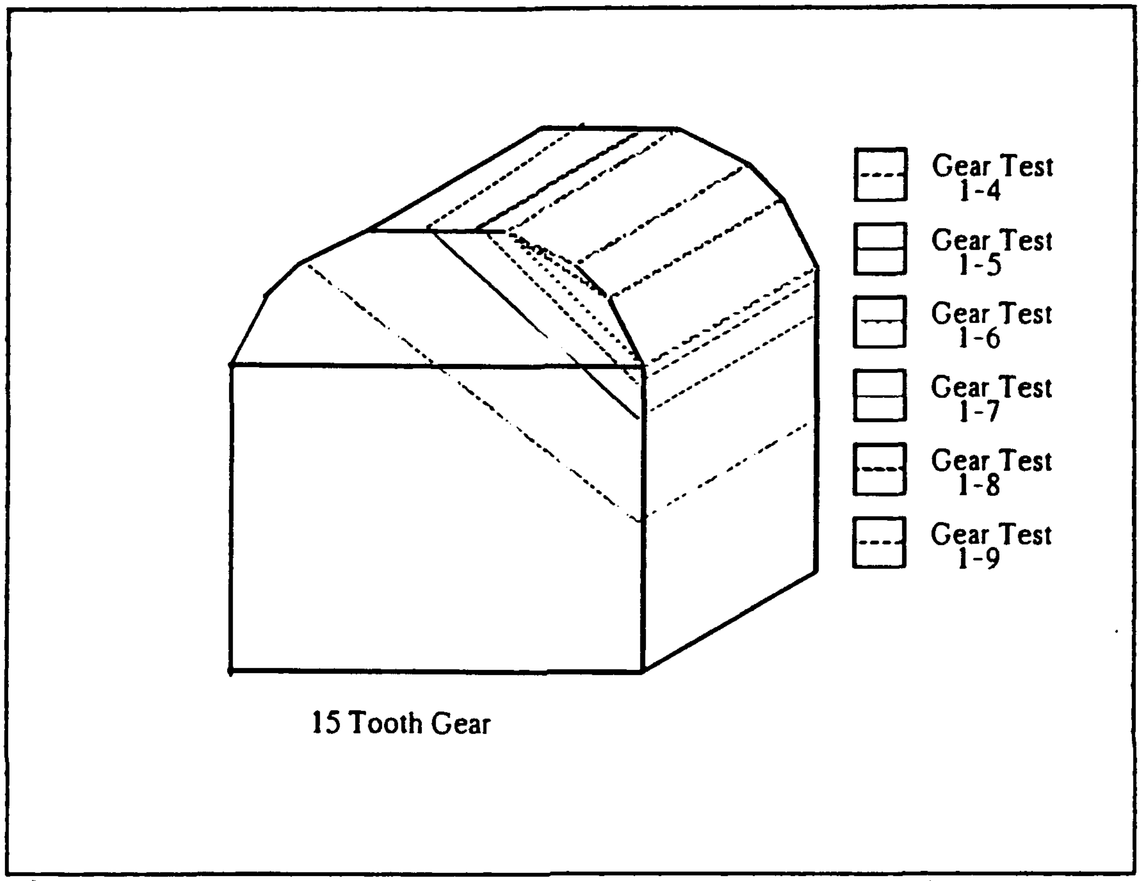

Figure 24 Gear Damage Levels for Gear Tests 1-4 through 1-9

b. Presentation and Discussion of Test Data

In general the damage to the 15 tooth pinion manifested itself through an overall reduction in gear mesh frequency amplitudes and increases in the 30 Hz sidebands. Additionally, as damage became severe, overall vibration levels increased throughout the frequency spectrum, being principally noted in the drive shaft rotative frequency and its harmonics. While all of these characteristics were

expected, there were several instances where sideband growth was lower in cases with a higher degree of damage than in cases involving lesser degrees of damage. This phenomenon can be partly explained by considering that the degree of contact between the damaged tooth and the mating gear tended to be considerably reduced as more material was removed thereby reducing the degree of impact. In the most extreme case, the damaged tooth may not have actually made contact at all, with the vibration increases experienced stemming from the misalignment experienced by the following tooth as it meshed with the mating gear.

Figure 26 illustrates a portion of a time signature from Gear Test 1-2. The 33 ms pulse stemming from the damaged tooth impacting as it goes through the gear mesh is predominant. Observation of the frequency spectrum in Figures 25 and 28 also reveals the strong influence of the 30 Hz sidebands throughout the spectrum but in particular about the gear mesh frequencies. Figure 27 presents the broad band Cepstrum. Here the predominant 33.3 ms quefrency and its rahmonics are clearly visible.

A summary of the means and standard deviations of the decibel levels extracted from the light damage level tests is provided in Table V along with the baseline values. A quick perusal of this data reveals that the most prominent deviations from the baseline occurred at 92, 450, 900, and 1350 Hz as well as at the 33.3 ms and 111 ms quefrencies.

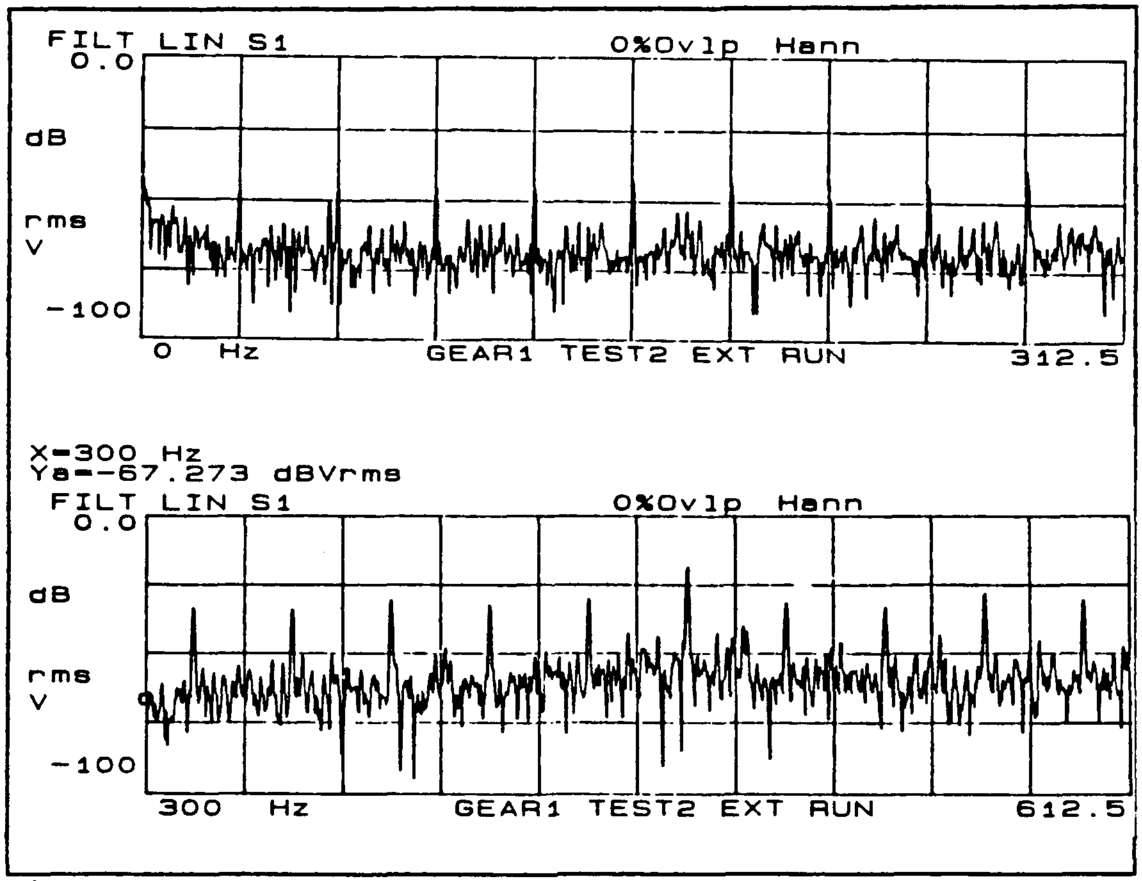

Figure 25 Frequency Spectrum for Gear Test 1-2 0-612 Hz

Additionally significant changes can be noted in the 30 Hz sidebands associated with the 900 and 1350 Hz gear mesh frequencies. Whereas the 30 Hz sidebands experienced significant growth, the gear mesh frequencies increased in magnitude on one occasion and decreased at the remaining frequencies that changed. Additionally there was an increase in the magnitude of the cepstrum at 33.3 ms and its rahmonics which was balanced by a drop in the magnitude at the 111 ms quefrency as well as at the bulk of the remaining cepstral quefrencies monitored.

Table V. Mean and Standard Deviations of dB Levels in Gear 1 Low Severity Fault Tests

Frequency (Hz)	9.0	18.0	30.0	60.0	92.0	103	118	184	236
Baseline	-61.2	-64.6	-57.6	-53.7	-55.0	-74.4	-70.8	-41.2	-64.4
Gear Test 1-4	-60.9 / 1.9	-65.8 / 2.4	-56.6 / 1.9	-51.4 / 0.9	-69.5* / 2.1	-74.2 / 0.9	-70.3* / 1.1	-43.6* / 1.9	-70.4 / 0.7
Gear Test 1-5	-61.3 / 2.5	-64.7 / 2.3	-56.5 / 2.9	-51.4 / 1.8	-68.1* / 2.3	-73.5 / 1.5	-69.6* / 3.2	-44.6* / 2.6	-65.0* / 2.1

Frequency (Hz)	450	9SB	30SB	900	9SB	30SB	1350	9SB	30SB
Baseline	-17.4	-45.5	-40.2	-18.1	-39.6	-37.4	-19.1	-34.1	-32.8
Gear Test 1-4	-17.6* / 2.1	-42.9 / 2.3	-41.9 / 1.4	-15.9 / 1.5	-37.0 / 1.6	-31.4 / 1.4	-23.3* / 2.9	-35.8 / 3.0	-29.0 / 2.6
Gear Test 1-5	-21.5 / 1.5	-45.6 / 2.6	-40.1 / 1.7	-14.8 / 0.5	-38.5 / 1.2	-27.3 / 1.7	-15.9 / 0.8	-33.6 / 1.2	-27.4 / 1.0

Cepstrum (ms)	8.5	9.7	10.9	10.9Av	33.3	33.3Av	111
Baseline	-6.4	-7.7	-3.7	-5.4	-4.4	-4.4	-6.7
Gear Test 1-4	-7.9 / 0.8	-6.2 / 0.5	-6.0 / 0.5	-6.1 / 0.5	-3.3 / 0.7	-4.0 / 0.3	-13.1 / 1.3
Gear Test 1-5	-6.2 / 0.8	-7.0 / 0.9	-6.9 / 0.6	-6.4* / 1.0	-2.3 / 0.6	-3.1 / 0.4	-12.1 / 0.7

* 1 outlier removed in computation of standard deviation and mean.

This phenomenon of an increase in the dB level in one region of the spectrum accompanied by a decrease in other regions, is often observed in the data presented, especially in cases of low to moderate damage to a component. However, this phenomenon is even more noticeable in the cepstrum. Since the vibration signature of a machine is analogous to an energy distribution, it should be expected that the overall spectrum

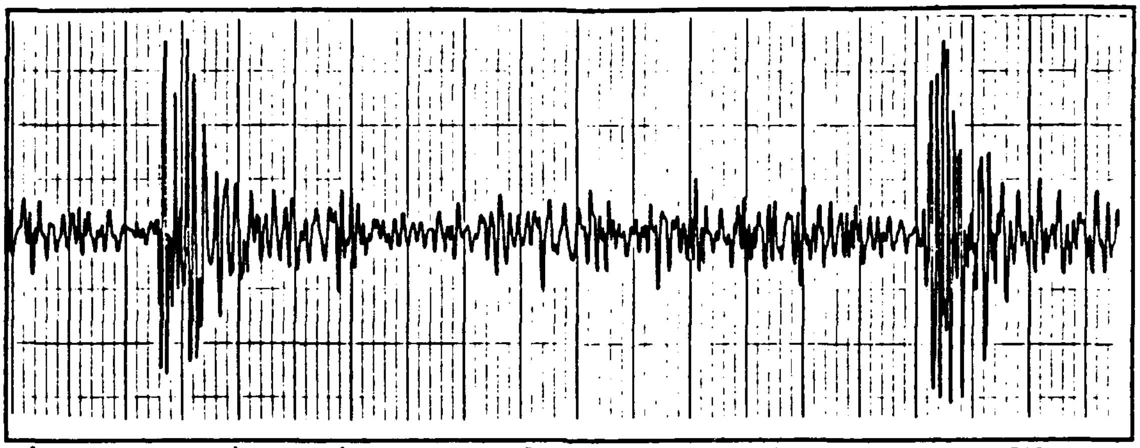

Figure 26 Time Signature for Gear Test 1-2; 5ms,5V per Division

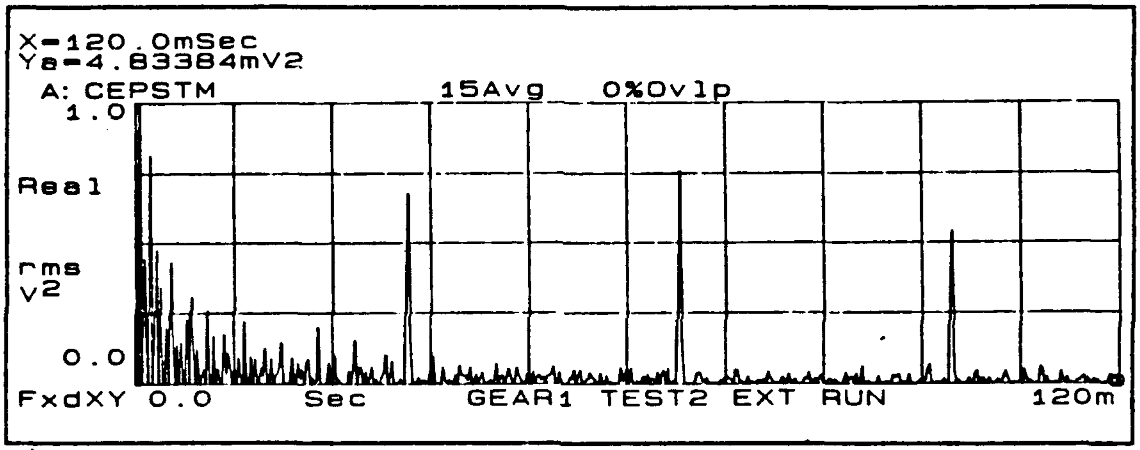

Figure 27 Cepstrum for Gear Test 1-2

possesses a finite amount of energy. Consequently an increase in energy at one frequency or family of frequencies should be expected to be accompanied by a decrease somewhere else. Furthermore, the location in the frequency spectrum where the energy level drops can be as significant for diagnostics purposes as the location where the energy rises. As additional empirical data is presented, it should be possible to identify the frequencies where this is the case.

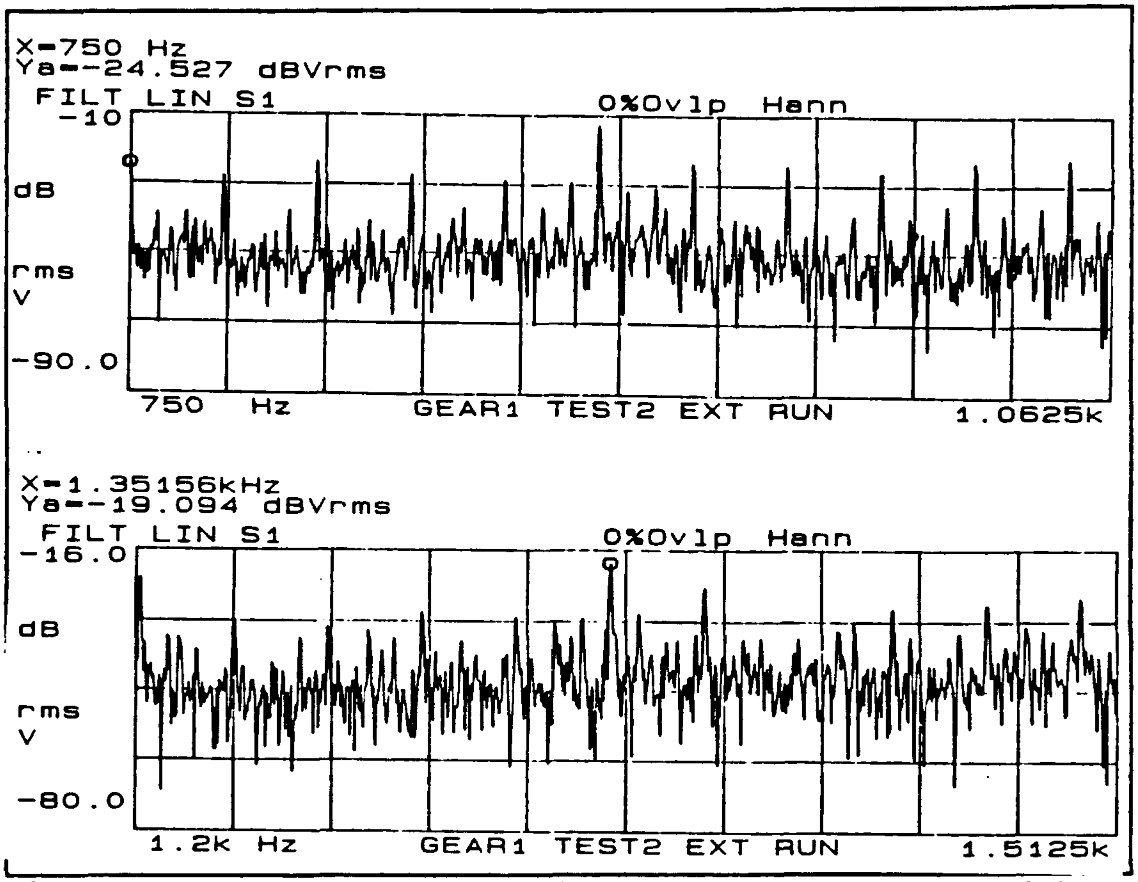

Figure 28 Frequency Spectrum for Gear Test 1-2 750-1512 Hz

Table VI presents the means and standard deviations for the dB levels encountered in the tests involving moderate damage to Gear 1. Upon observation of these results the following changes to the vibration signature are noted. The most prominent region of amplitude growth is consistently within the 30 Hz sidebands and the 33.3 ms quefrency associated with the 30 Hz sidebands. Additionally the 30 Hz shaft rotative frequencies experience a slight increase in excitation. The magnitude of the signals at the gear mesh frequencies alternately increase and decrease from test to test as do a number of the bearing frequencies. Since the

gear mesh frequencies have a direct connection to the diagnosis of gear faults, which are the faults being studied, it would appear that both positive and negative deviation of these dB values from the baseline are significant.

Gear Tests 1-2, 1-8, and 1-9 involved severe gear tooth damage and a summary of the data obtained from these tests is provided in Table VII. Gear Tests 1-2 and 1-8 reflect a continuation of the trend established in lower severity fault tests. However in Gear Test 1-2, there is an increase in vibration level at the shaft rotative frequencies and a number of the bearing frequencies in addition to the 30 Hz sideband and quefrency increases. This infers an overall increase in system energy which would appear to be characteristic to high severity faults. It is expected that at this point broad band vibration indicators sensing peak or RMS levels would register a significant fault.

Gear Test 1-9 had to be curtailed after only an incomplete set of readings had been taken due to a catastrophic failure. In this failure, the set screw affixing the 15 tooth pinion (Gear 1) worked itself loose and then moved down the shaft, ultimately binding with the 50 tooth gear (Gear 2) on one side. Damage to Gear 1 involved severe deformation of all teeth along at least 50% of the contact length of the gear. Damage to Gear 2 was considerably more mild, involving lesser deformations along the edge of the tooth, extending in the worst case to 25% of the contact

Table VI. dB Level Mean and Standard Deviations for Gear 1 Moderate Severity Fault Tests

Frequency (Hz)	9.0	18.0	30.0	60.0	92.0	103	118	184	236
Baseline	-61.2	-64.6	-57.6	-52.7	-55..0	-74.4	-70.8	-41.2	-64.4
Gear Test 1-1	-63.7 2.6	-63.6 2.1	-53.6 5.3	-50.2 2.6	-49.9 2.4	-70.9 1.4	-70.0 2.1	-38.7 5.6	-59.6 4.4
Gear Test 1-3	-59.6 2.2	-65.5* 2.3	-57.1 1.9	-51.5 1.0	-63.9 3.4	-72.9 2.1	-72.7 1.9	-39.8 2.1	-68.4 3.3
Gear Test 1-6	-60.1* 1.4	-62.1 2.0	-55.8 1.1	-53.9 1.8	-62.9 3.6	-73.6 2.0	-70.8 2.9	-57.3* 2.2	-68.3 2.4
Gear Test 1-7	-61.5 2.7	-64.5 1.1	-56.1* 2.2	-52.0 1.5	-60.3 1.9	-74.7 1.4	-66.0* 3.5	-56.6* 3.0	-68.1 1.8

Frequency (Hz)	450	9SB	30SB	900	9SB	30SB	1350	9SB	30SB
Baseline	-17.4	-45.5	-40.2	-18.1	-39.6	-37.4	-19.1	-34.1	-32.8
Gear Test 1-1	-13.2 0.7	-44.4* 1.5	-28.7 1.3	-23.4 2.7	-39.9 1.1	-29.1 2.3	-18.3 2.6	-34.9 1.9	-24.4* 1.9
Gear Test 1-3	-14.4 1.5	-39.6 1.8	-40.6 2.7	-16.7 1.6	-35.6 0.7	-32.0 1.3	-16.1* 1.4	-28.8 1.4	-26.5 1.1
Gear Test 1-6	-21.4 1.4	-41.9 2.5	-40.9 1.9	-18.6 2.0	-34.4* 3.7	-32.1 2.5	-22.7 2.9	-31.7 2.5	-29.7 2.7
Gear Test 1-7	-26.4 1.4	-46.7 1.2	-40.9 1.9	-16.3 0.6	-34.2 1.2	-28.8 1.8	-19.1 1.6	-31.3 0.9	-28.0 1.0

Cepstrum (ms)	8.5	9.7	10.9	10.9Av	33.3	33.3Av	111
Baseline	-6.4	-7.7	-3.7	-5.4	-4.4	-4.4	-6.7
Gear Test 1-1	-6.7 1.1	-8.0 0.2	-6.4 0.5	-6.0 0.4	-2.3 0.8	-1.9 0.3	-12.1 1.5
Gear Test 1-3	-9.3 1.4	-6.1 1.0	-5.1 0.5	-5.5 0.7	-4.4* 0.8	-3.9 0.4	-7.0 0.8
Gear Test 1-6	-6.9 0.7	-6.2 0.6	-7.2 0.7	-6.8 0.9	-5.2* 0.7	-5.5* 0.4	-12.0 0.9
Gear Test 1-7	-4.6 0.3	-6.5 0.7	-7.9 0.4	-6.9 0.3	-2.7 0.5	-3.6 0.3	-14.4 1.0

length. The damage to Gear 1 was classified severe, while the damage to Gear 2 was classified as moderate. The readings in Gear Test 1-9 were taken immediately prior to the casualty.

Here there is a massive increase in energy level throughout the spectrum, indicating a very severe fault was in progress.

Following this casualty, once all other components had been inspected for damage, a test was conducted with both damaged gears in place. The results from this test are summarized in Table VII. Here significant increases in vibration amplitude throughout the spectrum, including the 9.0 Hz sidebands, which had remained inactive until Gear Test 1-9. The only frequency components that dropped was the gear mesh frequencies, which dropped to levels never descended to before. Nevertheless, the highest increase in dB level occurred in the 9.0 Hz sidebands, revealing their higher level of damage. Oddly, Cepstral readings experienced an overall decrease in magnitude and apparently did not register the fault.

3. Faults to the Driven Gear

The set of tests involving the 50 tooth gear (Gear2) consisted of a total of four tests. In the first test, the 50 tooth gear that was subject to the casualty described in the previous section was operated with an intact drive pinion. This test was designated Gear Test 2-1 and the gear was considered to have suffered moderate damage. The next test involved a separate gear that had one tooth that had most of its material removed except immediately about its base. This test was designated Gear Test 2-2 and was considered to

Table VII. dB Level Mean and Standard Deviations for Test Involving Severe Damage to Gear 1 and Moderate Damage to Gear 2

Frequency (Hz)	9.0	18.0	30.0	60.0	92.0	103	118	184	236
Baseline	-61.2	-64.6	-57.6	-52.7	-55.0	-74.4	-70.8	-41.2	-64.4
Gear Test 1-2	-58.0 2.1	-58.2 2.8	-50.5 * 2.1	-50.6 * 4.5	-50.9 * 4.8	-70.6 2.9	-59.9 * 2.9	-40.4 3.4	-62.3 2.6
Gear Test 1-8	-60.4 2.3	-65.3 2.7	-54.7 1.3	-52.0 1.7	-55.7 1.2	-74.5 1.3	-67.6 3.0	-51.4* 3.6	-68.8* 2.7
Gear Test 1-9	-44.2 (-)	-50.2 (-)	-33.8 (-)	-36.2 (-)	-47.6 (-)	-69.4 (-)	-44.7 (-)	-42.5 (-)	-36.6 (-)
Gear Test 1-10	-57.9 * 2.9	-60.2 * 2.4	-50.6 * 4.1	-43.4* 2.7	-44.8 * 3.5	-72.1 1.2	-60.7 2.4	-36.9 1.9	-57.5 3.7

Frequency (Hz)	450	9SB	30SB	900	9SB	30SB	1350	9SB	30SB
Baseline	-17.4	-45.5	-40.2	-18.1	-39.6	-37.4	-19.1	-34.1	-32.8
Gear Test 1-2	-17.4 2.4	-45.2 1.6	-23.0 2.1	-13.1 0.9	-37.3 1.8	-26.4 1.3	-19.1 0.7	-34.5 0.9	-26.6 2.6
Gear Test 1-8	-26.6 3.3	-46.7 1.0	-36.9 2.1	-16.9 0.6	-35.7 1.4	-30.2 1.4	-22.7 1.3	-34.7 1.5	-29.9 0.7
Gear Test 1-9	-17.4 (-)	-35.1 (-)	-28.9 (-)	-17.4 (-)	-33.7 (-)	-28.3 (-)	(-)	(-)	(-)
Gear Test 1-10	-18.8 * 1.7	-38.5 2.2	-32.1 1.5	-25.6 * 2.9	-38.1* 1.8	-29.2 * 2.5	-26.4 * 2.8	-31.4 1.9	-31.0 1.2

Cepstrum (ms)	8.5	9.7	10.9	10.9Av	33.3	33.3Av	111
Baseline	-6.4	-7.7	-3.7	-5.4	-4.4	-4.4	-6.7
Gear Test 1-2	-5.9 0.4	-7.5 0.6	-8.2 1.1	-6.3 0.5	-2.4 0.6	-2.4 0.3	-12.3 0.9
Gear Test 1-8	-7.6 0.5	-7.1 0.4	-7.3 0.8	-6.3 0.4	-2.3 0.2	-3.2 0.2	-12.5 0.8
Gear Test 1-9	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Gear Test 1-10	-7.4 1.0	-7.3 0.3	-7.7 0.5	-7.4 0.4	-4.7 0.8	-4.6 0.7	-13.2 1.4

* 1 outlier removed during computation of mean and standadrd deviation.
(-) incomplete data due to interruption by casualty.

involve severe damage. Gear Test 2-3 was conducted with a previously undamaged gear where the engaged face of a gear tooth was filed down until the involute shape was just barely affected. This level of damage, while considered of low

severity, produced an audible clicking sound which was also heard in the previous two tests involving Gear 2 damage. The last test of the series, Gear Test 2-4 involved expanding the damage imposed in Gear Test 2-3, removing the face and upper land on the engaged side but not affecting that of the disengaged face. The level of damage imposed was regarded as moderate. A schematic of the damage imposed in these tests is provided in Figure 29.

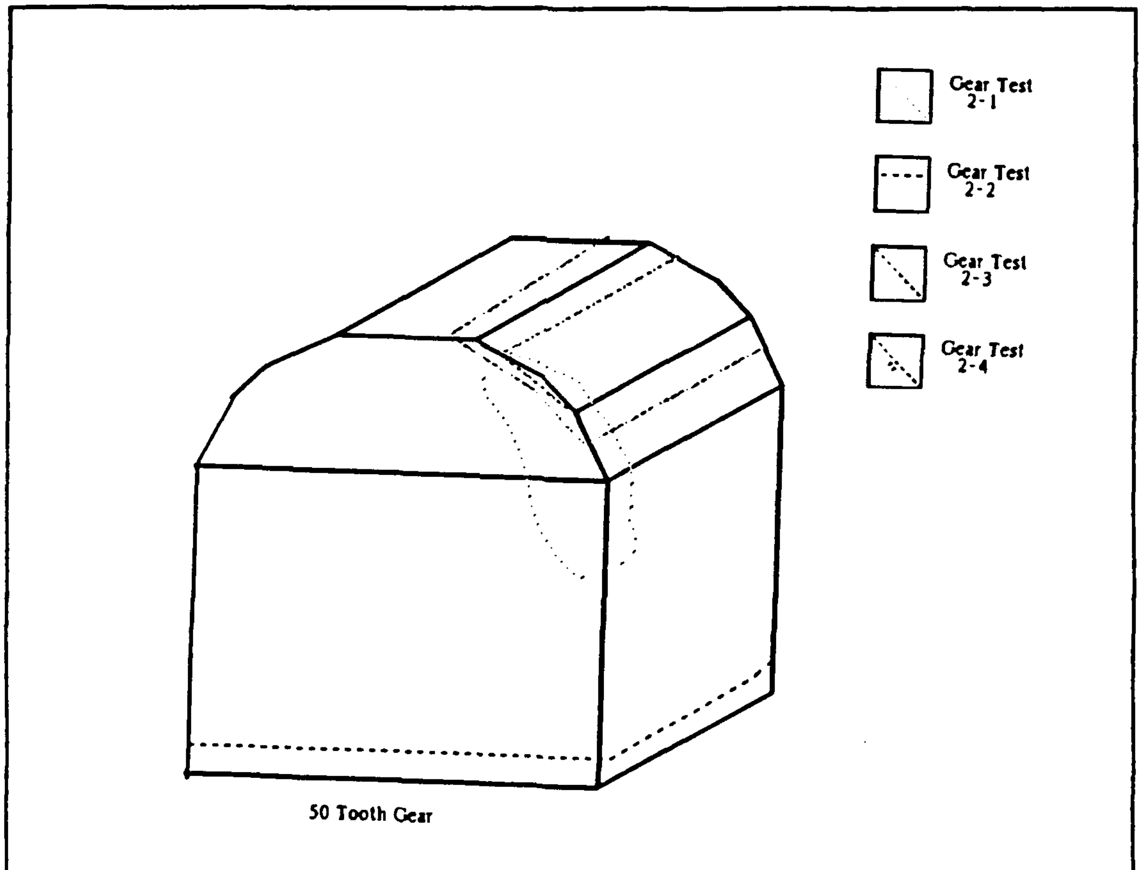

Figure 29 Damage imposed on Gear 2

Representative frequency spectra and broad band cepstral plots are provided in Figures 30 through 32. In these plots the 9.0 Hz sidebands and 111 ms cepstrum predominate as

is expected from the nature of the faults. Additionally the representative time domain plot in Figure 33 reveals sharp impacts occurring at a period of 110 ms, also corresponding to the Gear 2 rotative frequency.

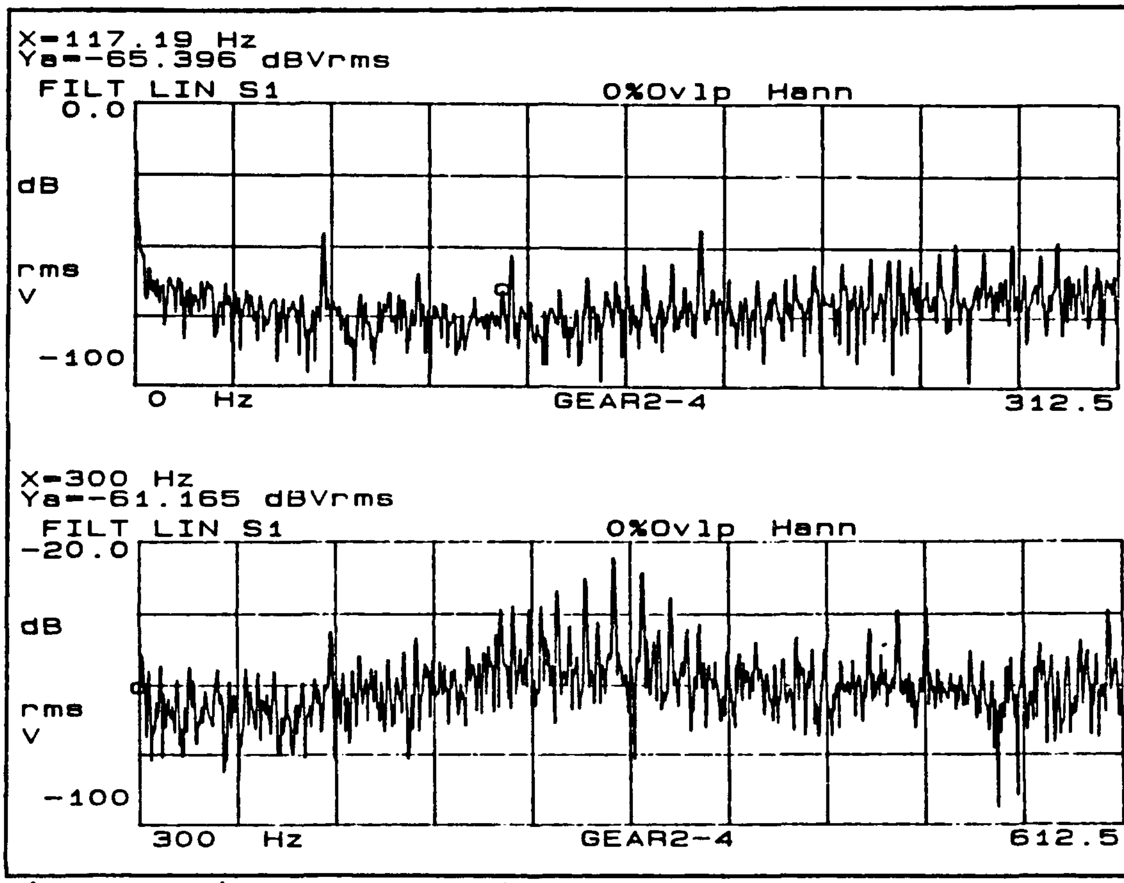

Figure 30 Linear Spectrum for Gear Fault 2-4; 0-612 Hz

Table VIII provides a summary of the dB levels experienced for the frequencies and quefrencies monitored. A brief inspection of the data will reveal the following trends. Observation of the averaged sidebands for the machine clearly indicates a fault in Gear 2 even in the case of low severity damage. The fault appears to become evident

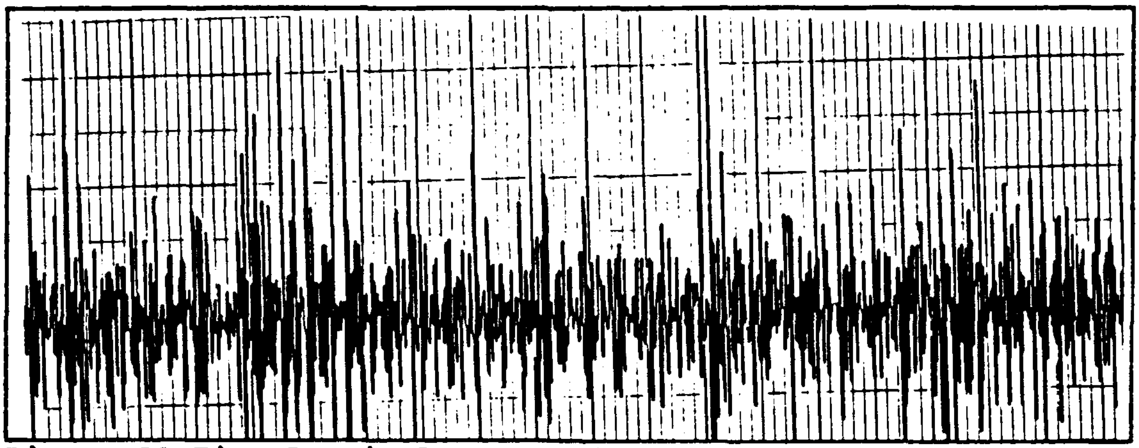

Figure 31 Time Domain Response for Gear Test 2-4; 5V/5ms per Division

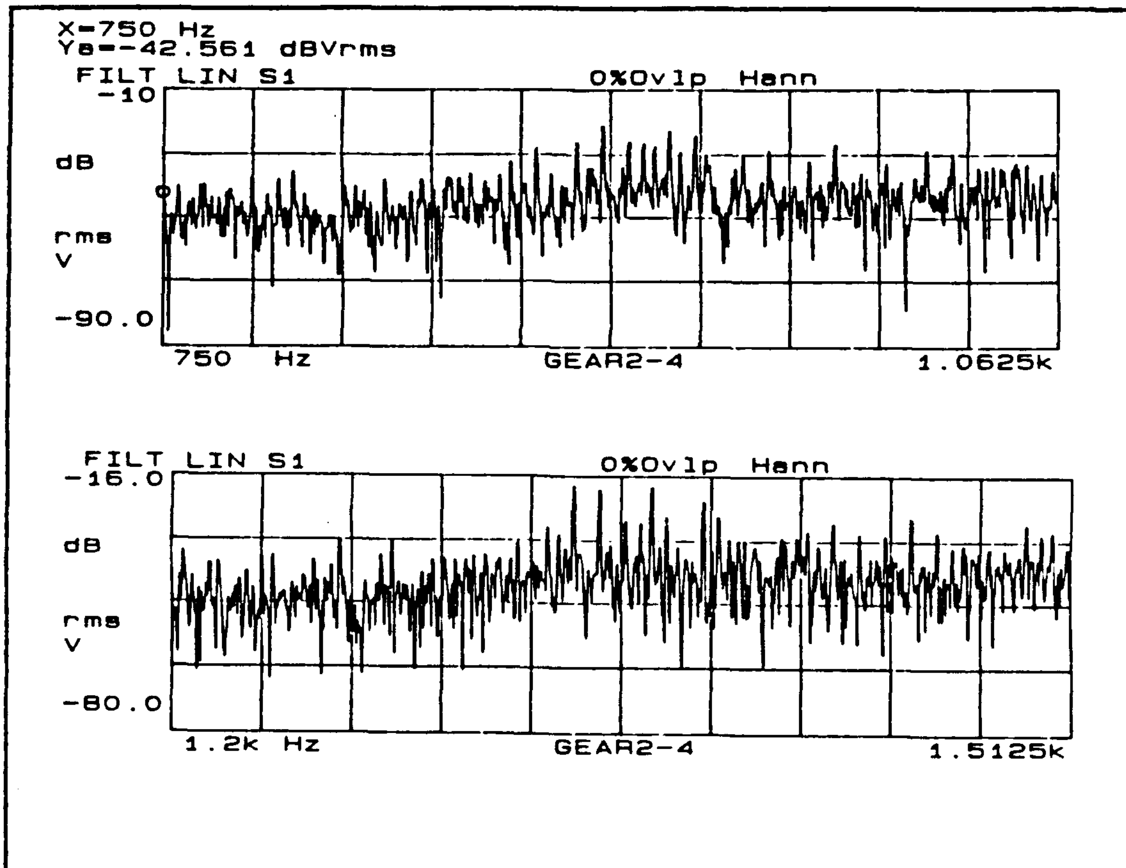

Figure 32 Linear Spectrum for Gear Fault 2-4; 750-1512 Hz

first in the sidebands about the 450 and 900 Hz gear mesh frequencies. The sidebands about 1350 Hz appear to undergo a

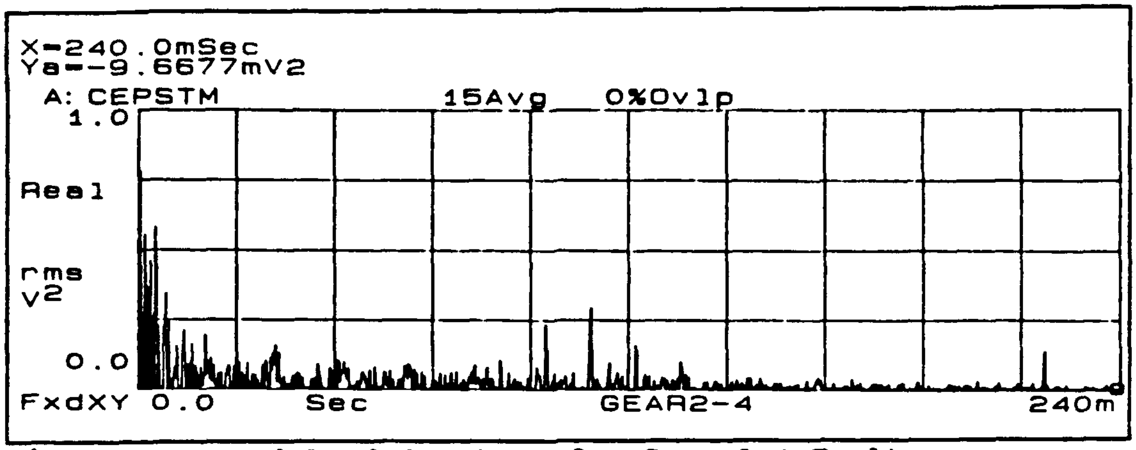

Figure 33 Broad Band Cepstrum for Gear 2-4 Fault

lesser degree of excitation which actually declines in the highest severity faults.

Gear mesh frequencies again predominantly experience dB drops in all but the most severe cases. Surprisingly, in all but the most severe cases, the 9.0 Hz shaft frequencies remained relatively unaffected in all but the highest level of damage, even though they would appear to be most directly coupled to the damaged gear. Conversely, in all cases the 30 Hz shaft frequencies, which appear to be relatively remote from the damage, underwent large dB rises.

In moderate and high severity faults both the bearing inner race frequencies and their related quefrencies experienced some increase in vibrational amplitude. However, with the possible exception of the sidebands, the most bold indication of gear damage consistently was the 111 ms cepstrum.

Table VIII. Mean and Standard Deviations for dB levels for Gear 2 Fault Tests

Frequency (Hz)	9.0	18.0	30.0	60.0	92.0	103	118	184	236
Baseline 1	-61.2	-64.6	-57.6	-52.7	-55.0	-74.4	-70.8	-41.2	-64.4
Gear Test 2-1	-63.1 2.7	-64.8 3.5	-62.6 2.1	-49.2 0.4	-57.7• 1.2	-74.3 1.2	-70.9 • 3.5	-43.5•• 3.0	-68.4 • 4.8
Gear Test 2-2	-53.7• 3.5	-55.6• 4.2	-57.6• 3.6	-48.3 0.8	-53.6• 5.3	-74.8 1.3	-54.2 • 5.6	-46.6• 3.6	-59.9 • 3.9
Baseline 2	-61.8	-63.9	-67.2	-48.3	-69.4	-74.4	-74.0	-52.5	-66.9
Gear Test 2-3	-62.6• 2.6	-66.3 2.2	-63.0 1.8	-46.7 0.5	-62.4 2.3	-72.1 1.4	-68.6 5.6	-48.5 1.8	-65.3 2.6
Gear Test 2-4	-62.5 3.3	-65.6 2.3	-65.3 2.6	-45.2 0.6	-61.5 2.8	-74.0 1.3	-70.0• 2.6	-48.9 3.8	-68.0• 3.2

Frequency (Hz)	450	9SB	30SB	900	9SB	30SB	1350	9SB	30SB
Baseline 1	-17.4	-45.5	-40.2	-18.1	-39.6	-37.4	-19.1	-34.1	-32.8
Gear Test 2-1	-20.6 • 1.6	-40.5 2.7	-38.7 2.9	-23.1 2.0	-37.9 2.2	-38.6 1.7	-23.4 • 3.2	-32.6 • 2.4	-37.8 2.2
Gear Test 2-2	-22.5• 2.8	-37.5 1.3	-39.8 1.0	-21.6 2.2	-34.8 2.3	-43.5 1.2	-22.4• 2.7	-34.0 1.6	-36.2 1.3
Baseline 2	-16.3	-33.4	-41.0	-17.1	-32.3	-39.9	-25.4	-30.4	-34.0
Gear Test 2-3	-23.1 3.0	-41.9 6.0	-50.0 2.3	-23.0 1.1	-31.8 1.6	-38.7 1.2	-24.5 2.7	-30.1 0.8	-37.0 -1.5
Gear Test 2-4	-23.5 1.0	-36.4 0.8	-50.0 2.3	-26.3 3.2	-30.7 1.0	-38.3 3.3	-27.3 2.3	-24.8 1.5	-34.0 1.9

Cepstrum (ms)	8.5	9.7	10.9	10.9Av	33.3	33.3Av	111
Baseline 1	-6.4	-7.7	-3.7	-5.4	-4.4	-4.4	-6.7
Gear Test 2-1	-5.0 0.9	-6.6 1.0	-5.1 0.6	-6.5 0.3	-6.5 0.6	-5.8 0.5	-5.8• 1.4
Gear Test 2-2	-6.0 0.5	-7.9 0.3	-6.6 0.6	-8.3 0.6	-8.0 0.6	-8.3 0.5	-3.0 0.7
Baseline 2	-5.3	-7.7	-6.0	-7.8	-7.6	-7.8	-10.9
Gear Test 2-3	-7.5 0.6	-6.8 0.4	-7.4 0.8	-8.2 0.5	-7.8 0.6	-8.9 0.6	-7.0 0.5
Gear Test 2-4	-7.7 0.6	-7.7 0.6	-6.5 0.3	-8.4 0.3	-8.3 0.4	-8.2 0.4	-5.3 0.3

• One outlier removed for computation of mean and standard deviation.

4. Bearing Faults

Acquisition of bearing fault data was rather difficult to accomplish. The small size of the bearings limited the

degree of control on the severity and location of damage that could be imposed. Additionally, the bearings were very lightly loaded and, compared with the gear related signals, the bearing signals were barely recognizable from the ambient noise. Finally, a computational error was made which rendered two tests involving the low speed shaft bearings useless. Thus, the only data presented and utilized in the moderate complexity neural networks involves a high speed shaft bearing whose operation became increasingly rough due to a lack of lubrication. To make matters worse, these tests were conducted immediately after the wear-in period of both Gears 1 and 2 following the casualty experienced during Gear Test 1-9. As a result there was a high degree of gear noise from both gears.

On the other hand these tests appeared to be good examples of multiple component faults on which conventional rule based expert systems perform questionably and thereby were retained. Because of the continued wearing in of the new gears, Gear 1 was determined to have a "moderate" severity damage equivalent while Gear 2 was determined to posses a low severity damage equivalent. The poorly lubricated bearing was determined to posses a low severity damage level due to its size and loading. A summary of the results from these tests is provided in Table IX.

Investigation of this data immediately indicates that the prominent signal stems from the gears wearing in. However, there are significant increases in vibration magnitudes at 92

Table IX. dB Level Mean and Standard Deviation for Tests Involving Bearing Faults

Frequency (Hz)	9.0	18.0	30.0	60.0	92.0	103	118	184	236
Baseline 2	-61.8	-63.9	-67.2	-48.3	-69.4	-74.4	-74.0	-52.5	-66.9
Bearing 1-1	-62.0 2.3	-65.9 2.2	-64.8 1.3	-49.0 1.0	-60.7* 1.6	-67.6 0.3	-73.8 1.1	-51.4* 3.4	-60.8* 2.6
Bearing 1-2	-62.4 3.9	-65.5 3.6	-65.2 1.4	-48.8 0.5	-63.8 1.8	-70.1 4.3	-74.2 1.9	-55.2 2.6	-58.4 0.7

Frequency (Hz)	450	9SB	30SB	900	9SB	30SB	1350	9SB	30SB
Baseline 2	-16.3	-33.2	-41.0	-17.1	-32.3	-39.9	-25.4	-30.4	-34.0
Bearing 1-1	-8.3* 1.9	-31.4 2.2	-32.2* 3.0	-21.8 1.9	-37.4 2.4	-37.6* 3.4	-19.9* 3.6	-33.9 4.1	-36.0 4.0
Bearing 1-2	-7.3 1.1	-30.5 0.4	-35.7 0.7	-15.6 0.6	-32.2 1.4	-31.2 1.1	-18.2 0.5	-25.1 0.0	-29.3 1.6

Cepstrum (ms)	8.5	9.7	10.9	10.9Av	33.3	33.3Av	111
Baseline 2	-5.3	-7.7	-6.0	-7.8	-7.6	-7.8	-10.9
Bearing 1-1	-5.5 0.7	-8.0 0.2	-5.4 1.0	-6.1 0.6	-4.8 1.1	-5.5 0.8	-8.1 0.8
Bearing 1-2	-6.3 1.2	-8.0 0.4	-4.5 0.6	-6.3 0.0	-4.9 0.8	-5.8 0.1	-10.6 1.1

Hz, and 236 Hz, as well as in the 10.9 and 9.7 ms quefrencies. These correspond to the bearing inner and outer races as well as the balls themselves.

5. Shaft Faults

Shaft faults were imposed by two methods. In the first, a shaft imbalance was imposed by allowing the high speed shaft to operate unsupported by the remote bearing with respect to the motor coupling. This test, designated Shaft Test 1-1, while producing relatively low vibration levels, generated a highly visible imbalance and was therefore assigned a damage severity of "moderate". The second type of

shaft fault imposed involved replacing the slow speed shaft with one that was "slightly" bent. This misalignment generated both a highly visible wobble in the shaft and produced very strong vibrational signals. Two of these tests were conducted; one involved the use of a 15 tooth pinion whose teeth had suffered an excessive degree of generalized wear and was in need of replacement. The second test used a replacement gear that was in good condition. Accordingly, the first of these tests was assigned a low severity level for the gear, a high severity level for the 9.0 Hz shaft, and was designated Shaft Test 2-1. Similarly, the second test involved a damage severity rating of "severe" for the shaft, "normal" for the gear, and was designated Shaft Test 2-2.

Representative plots of the linear frequency spectrum and cepstrum are provided for Shaft Test 2-2 in Figures 34 through 36. The strong signal generated by the shaft is clearly visible in the 0-312 Hz frequency plot as are strong 9.0 Hz sidebands generated as Gear 2 alternately loads and unloads each shaft rotation. Additionally, a time domain plot illustrating the pulses generated by the bent shaft is provided in Figure 37.

A summary of test results is provided in Table X. A brief investigation of this data reveals the following. In Test Shaft 1-1, there was relatively little deviation from the baseline. There was a significant increase in the shaft rotative frequency and alternately increasing and decreasing

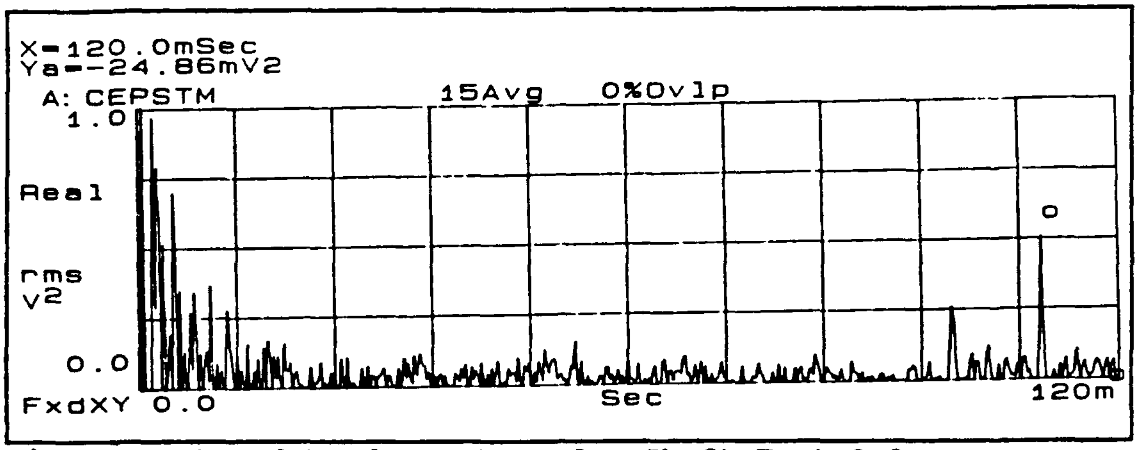

Figure 34 Broad Band Cepstrum for Shaft Test 2-2

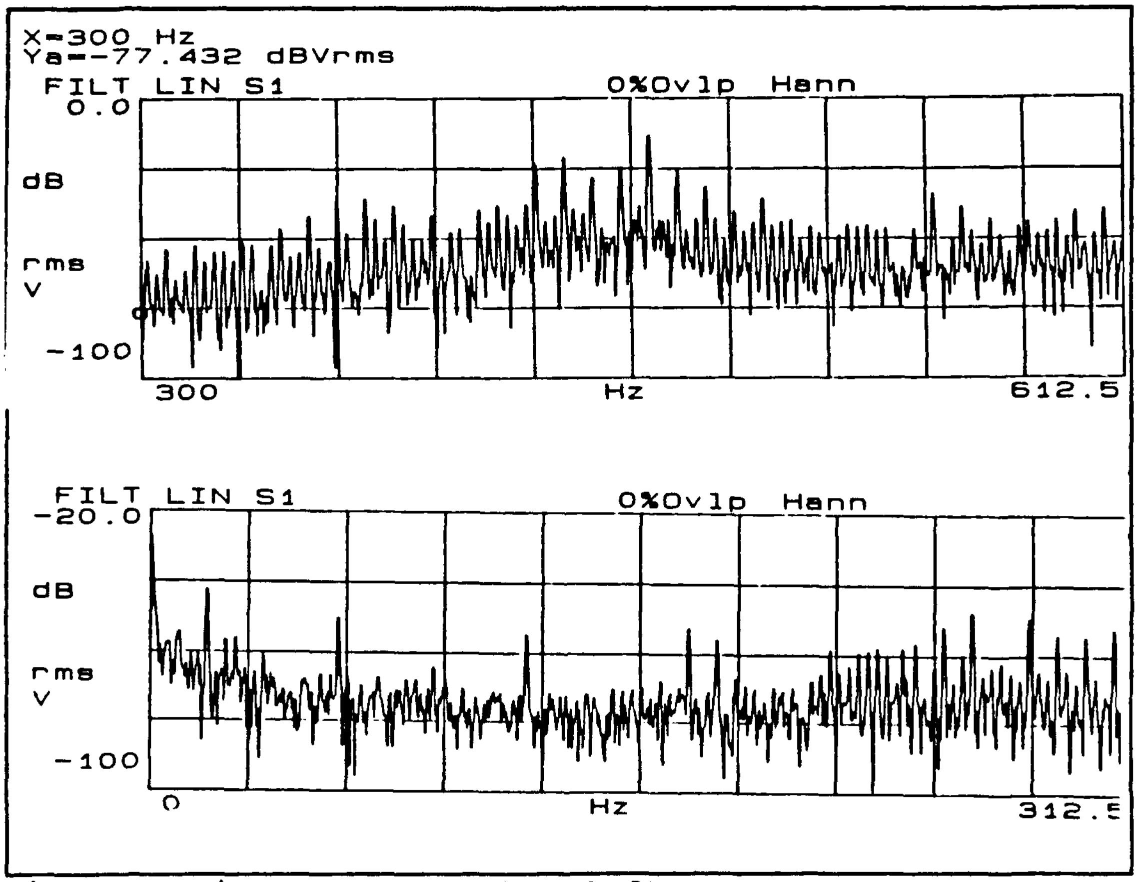

Figure 35 Linear Spectrum for Shaft Test 2-2; 0-612 Hz

dB levels at the gear mesh frequencies. There was a slight increase in the 30 Hz sidebands about 900 Hz and a significant

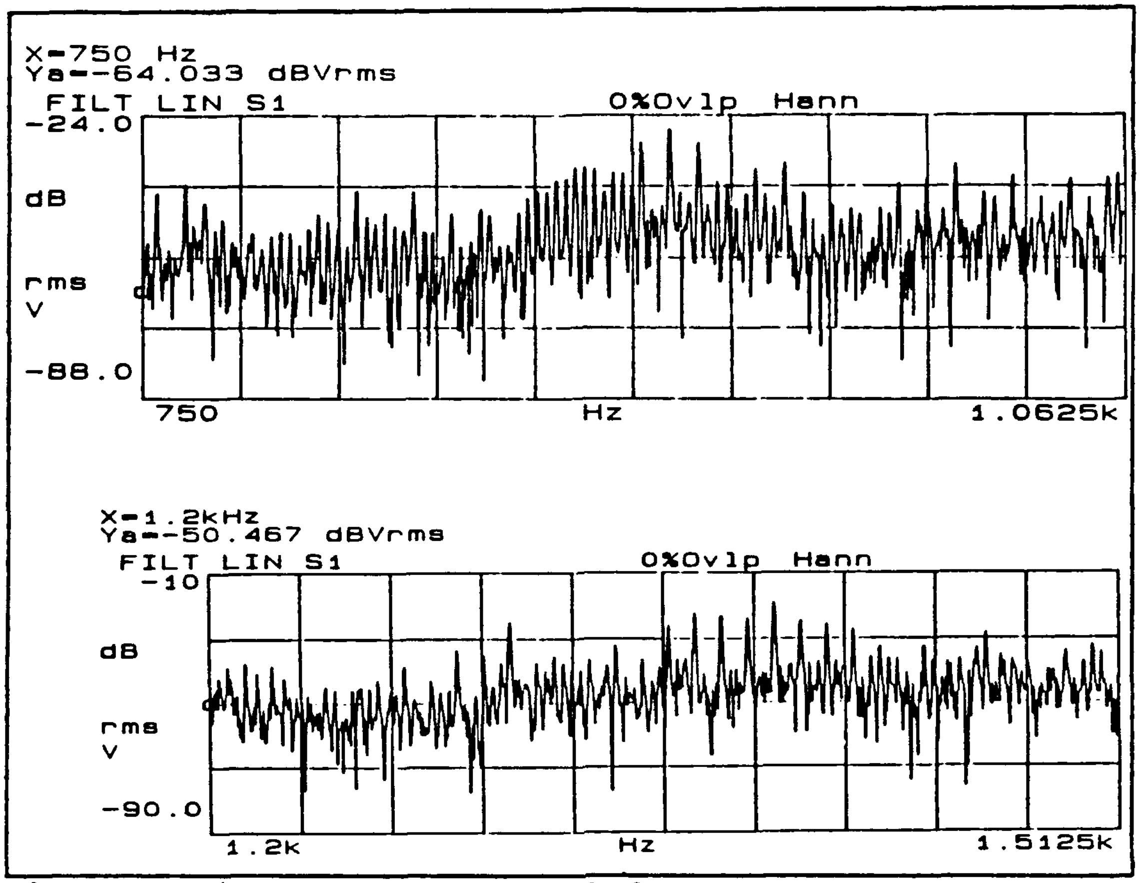

Figure 36 Linear Spectrum for Shaft Test 2-2; 750-1512 Hz

increase in the 33.3 ms cepstrum and the average of its rahmonics. The changes to the gear mesh frequency , sidebands, and cepstrum can be attributed to the 15 tooth pinion alternately loading and unloading as the shaft is allowed to deflect; The 30 Hz dB increase relates directly to the shaft imbalance.

Shaft Tests 2-1 and 2-2 varied considerably from Shaft Test 1-1. Both the shaft rotative frequency and even more noticeably its first harmonic have strong increases in magnitude. However, there are massive drops in dB at the gear

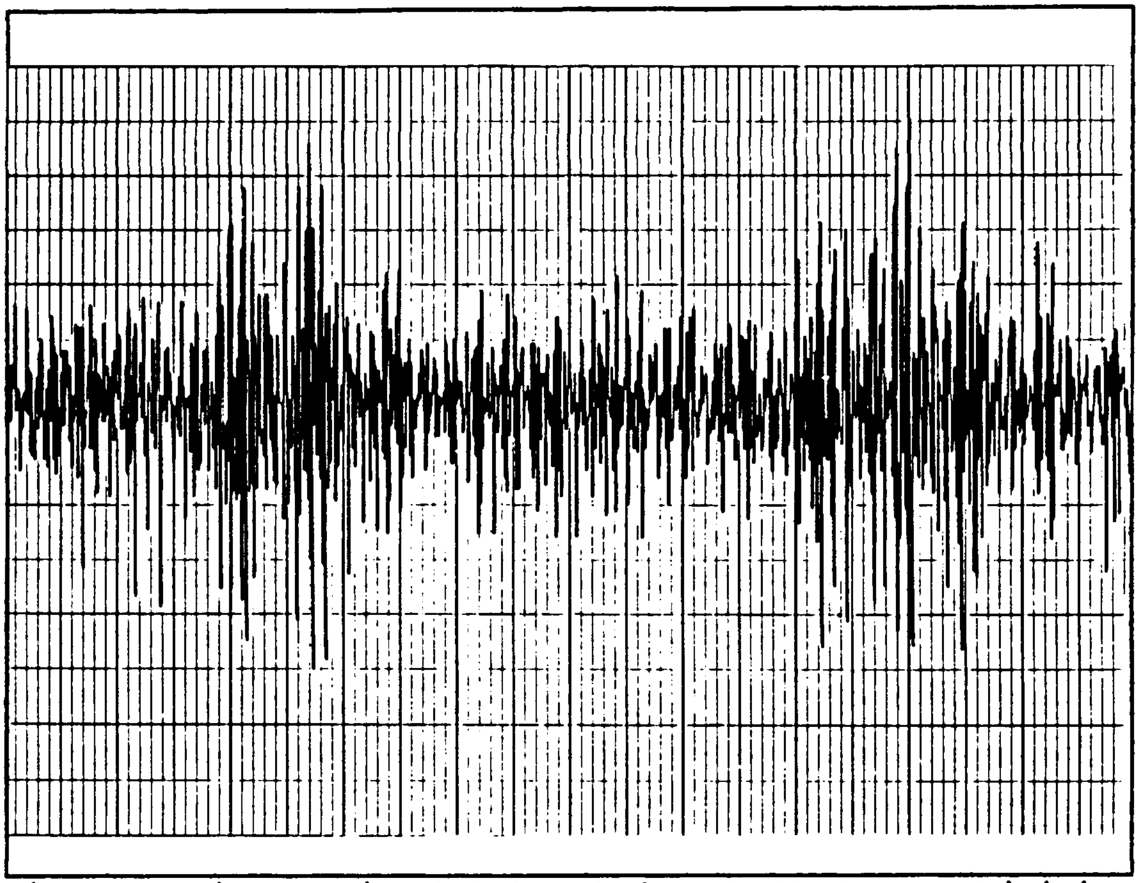

Figure 37 Time Domain Plot for Shaft 2-2 2V,20ms per Division

mesh frequencies and noticeable gains at the 9.0 Hz sidebands in both tests. While there are no significant gains in the cepstrum for the quefrencies monitored, there was a significant gain at 222 ms, a rahmonic of the 9.0 Hz family of frequencies. While the 9.0 Hz sideband growth in

Shaft Test 2-1 can be explained in part by the gear damage, the only explanation for this in Shaft Test 2-2 is the sinusoidal loading and unloading of the gear as the bent shaft rotates. Further, the dB levels in Shaft Test 2-2 are by and

Table X. Summary of Mean and Standard Deviations for Tests Involving Shaft Faults

Frequency (Hz)	9.0	18.0	30.0	60.0	92.0	103	118	184	236
Baseline 2	-61.8	-63.9	-67.2	-48.3	-69.4	-74.4	-74.0	-52.5	-66.9
Shaft Test 1-1	-59.5 1.8	-63.9 2.3	-65.0 1.6	-47.1 1.5	-69.8 2.7	-74.5 1.2	-72.7 1.7	-53.8 * 2.9	-65.7 * 1.6
Shaft Test 2-1	-55.0 2.1	-51.4 1.9	-64.3 1.9	-50.1 1.9	-69.2 3.3	-71.7 1.3	-72.6 1.4	-68.3* 1.8	-69.5 2.5
Shaft Test 2-2	-51.9* 2.3	-42.6 0.7	-64.8* 2.3	-49.6 1.2	-65.8* 2.6	-73.8 1.8	-72.3 1.7	-56.7 1.8	-71.0* 2.4

Frequency (Hz)	450	9SB	30SB	900	9SB	30SB	1350	9SB	30SB
Baseline 2	-16.3	-33.4	-41.0	-17.1	-32.3	-39.9	-25.4	-30.4	-34.0
Shaft Test 1-1	-21.3 1.8	-38.3 2.1	-44.4 1.8	-16.1 0.8	-34.5 0.7	-36.7 0.9	-17.9* 2.5	-25.2 1.4	-31.4 1.6
Shaft Test 2-1	-12.5 1.2	-25.4 1.5	-40.0 1.0	-30.6 3.9	-28.4 1.0	-43.7 2.0	-20.2 2.5	-29.0 2.0	-35.5 1.1
Shaft Test 2-2	-20.7 1.9	-24.1 0.4	-43.3 1.1	-29.8 1.9	-35.6 1.2	-44.4 1.3	-25.4 2.6	-25.5 2.4	-39.6 1.6

Cepstrum (ms)	8.5	9.7	10.9	10.9Av	33.3	33.3Av	111
Baseline 2	-5.3	-7.7	-6.0	-7.8	-7.6	-7.8	-10.9
Shaft Test 1-1	-5.7* 0.6	-7.0 1.1	-7.1 0.4	-7.8 0.4	-6.0 0.4	-6.0 0.3	-9.9 0.5
Shaft Test 2-1	-5.2 0.3	-7.4 0.6	-8.0 0.5	-8.7 0.5	-8.4 0.5	-8.2 0.3	-13.0 0.4
Shaft Test 2-2	-5.2 0.5	-6.0 1.0	-5.2 0.4	-8.2 0.6	-8.7 0.5	-8.6 0.3	-2.9 0.9

* One outlier removed from data set during computation of mean and standard deviation.

large greater than in Shaft Test 2-1, which runs counter to the conventional wisdom where higher damage levels yield higher magnitude vibration signals.

6. Summary of General Trends

In general, the following trends were observed as a result of the tests conducted on the physical model. First,

the faults imposed in most cases generated the type of vibration signatures that one would be led to expect from the elementary machinery condition monitoring and diagnostics practices discussed in Chapter III. However, there was a great deal more coupling between the various components than one may have expected, especially in cases involving the more severe damage levels. For example, in the case of the shaft faults to the low speed shaft, the vibration levels associated with the drive gear were considerably greater than those associated with the shaft itself. This could be accounted for by consideration of the small size of the model on which the tests were performed. Because of the small size of the model and light radial loads, bearing faults were particularly difficult to impose and detect. Nevertheless, analysis of the frequency spectrum and cepstrum did reveal bearing fault conditions to a limited degree in spite of the physical shortcomings of the model. Although at most frequencies monitored, dB decreases appear to have little relevance to the location of machinery faults, they do appear to be very significant in the case of gear mesh frequencies, where they tended to isolate the location of the fault to one of the two meshing gears. This observation would prove to be a key factor in the preprocessing of the vibration data prior to input into the neural networks.

VI. DIAGNOSTIC SYSTEM PROTOTYPE: THE NEURAL NETWORK

The neural network system designed to provide machinery diagnostics for the uncomplicated machinery described in the previous chapter was essentially an expansion of the simple diagnostics model described in Chapter IV. As there was still some question as to the relative effectiveness of the various frequencies and quefrencies monitored, particularly with respect to the isolation of gear faults by either sideband averaging or cepstral analysis, it was determined to develop two diagnostics neural networks; one utilizing sideband averaging about the first three gear mesh frequencies and the other utilizing cepstral inputs to supplement both gear mesh and bearing frequencies in the determination of gear and bearing faults. Additionally, both networks would receive shaft frequency inputs to aid in the diagnostics of shaft related faults.

All neural networks described in this section were created and trained on an IBM 386 personal computer utilizing Neuralware Inc.'s Neuralworks Professional II software simulator. Training sessions were limited to no more than one day run time, over which period a number in the order of 300,000 training presentations would occur.

Each of these networks were initially trained utilizing artificially generated data. This data was generated in the same manner as that of Chapter IV, but featured a different dB range to severity level correlation for each monitored parameter based on the statistics from the empirical baseline experiments. Then the networks were trained afresh using a portion of the data extracted from the tests described in the previous chapter. Finally, the networks trained on artificially generated data were first tested on a separate set of artificially generated data, whereas all networks under investigation were tested on a separate empirically based data set. As significant flaws were discovered in the performance of both basic networks when presented empirical data, a third diagnostic system utilizing both cepstral and sideband information was also investigated.

In this chapter the three rotative machinery diagnostics neural networks developed will be described. First, the general system architecture will be discussed, followed by a description of each network's inputs. Following this, the nature of the training sets and the preprocessing required will be described. Third, the results of the various tests and an evaluation of each network's performance will be presented. Finally, an evaluation of the relative effectiveness of the network inputs in each of the networks will be made.

A. SYSTEM ARCHITECTURE

Determining an effective system architecture is as important in solving practical engineering problems as is selecting a set of inputs to the network that adequately describes the decision space. Because this aspect of the problem is so important, a brief description of a preliminary architecture that was discarded in this particular application is as instructive as a description of the architecture eventually decided upon.

1. Preliminary Network Architecture

Originally the system architecture under consideration was patterned after the architecture utilized by Dietz, Kiech, and Ali[Ref.7] in their backpropagation diagnostics model for determining the location and severity of jet engine system faults. In this architecture, two levels of neural networks were used. The lower level determined the location of the fault and provided an input to the upper level network which noted the time duration of the fault signals to determine the severity of the fault. This two stage diagnostics system architecture was also used successfully by Watanabe and Himmelblau[Ref.1] in the detection of incipient faults in chemical processes.

The system architecture under consideration involved employing a series of pretrained severity indicating backpropagation modules similar to the simple diagnostics

model described in Chapter IV to provide a severity level ranging from 0.0 to 1.0 from each of the monitored parameters to an upper level neural network. As in the simple diagnostics model, these lower level networks were trained to classify a series of dB differences into "no", "low", "moderate" and "severe" fault conditions. The upper level network received these severity level inputs and identified the location of the fault by means of a binary output corresponding with each machinery component under scrutiny; "0" indicating no fault and "1" indicating a fault condition at that location. A schematic of this arrangement is provided in Figure 38.

A preliminary upper level network consisting of 18 inputs, 27 hidden, and 5 output PE's was successfully trained and tested. Additionally, as empirical data became available, the lower level networks were trained to provide severity indications for inputs that had severity criteria that departed from the uniform severity criteria established in Chapter IV. However, several lower level networks appeared to converge to a minimum level of RMS error but, when tested were found to produce grossly erroneous outputs. A probable cause of this anomaly was that the data set contained a large amount of zeros in both is input and desired output and the learning algorithm in place was a normalized cumulative delta rule which calculated RMS error over the entire epoch and averaged it. Because of the large number of low magnitude errors averaged with the large magnitude errors, the RMS error was

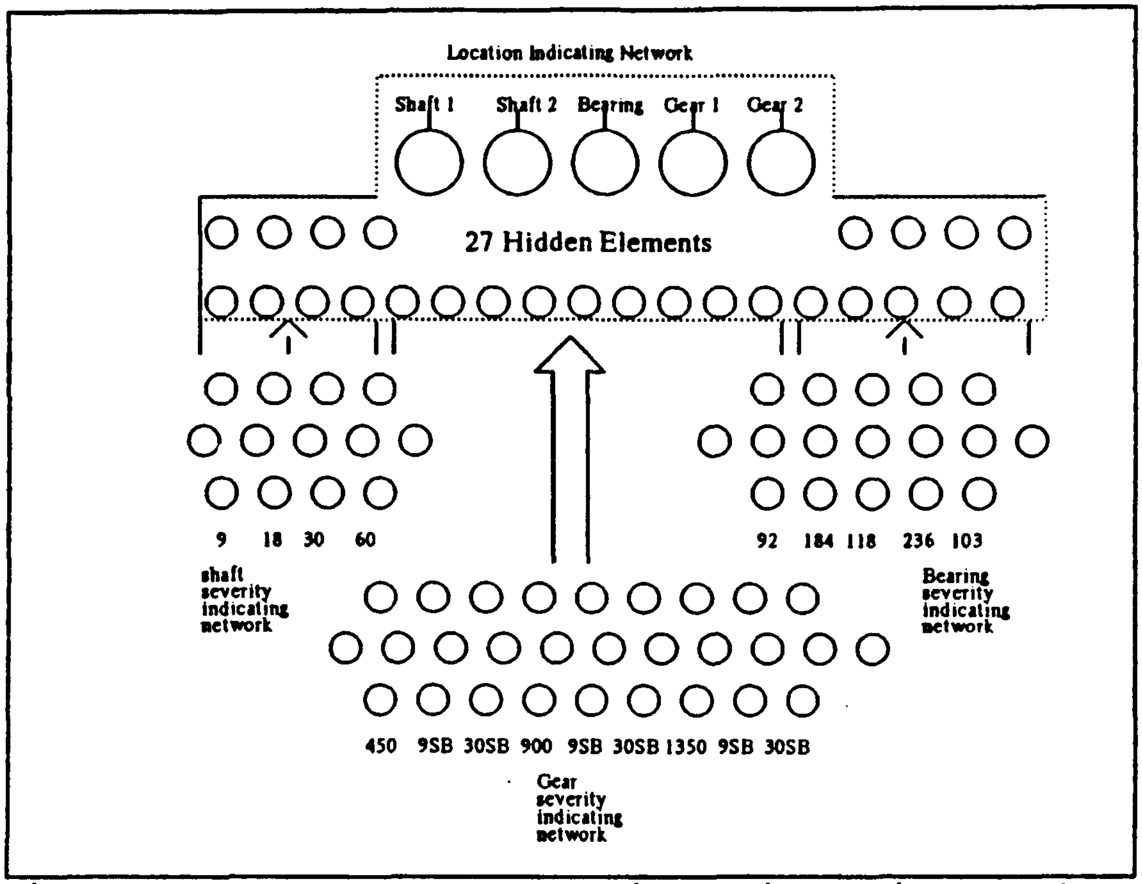

Figure 38 Proposed Two Level Machinery Diagnostic Network

misleadingly suppressed. Ultimately the cause of the gross errors themselves was attributed to inadvertently passing the input vectors through a linear mapping routine provided in the Neuralworks Professional II software simulator called a "MinMax Table". Essentially, this routine provided the network, which was tasked to provide a non-linear mapping of the inputs to values from 0.0 to 1.0 with an input already linearly mapped from 0.0 to 1.0, thereby making it very difficult to adjust weights effectively. By the time this cause was identified, however, an alternative architecture had been discovered and implemented with some degree of success.

Ironically this architecture featured a capitalization of that which proved to be the downfall of the originally proposed architecture, the MinMax Table.

2. Prototype Diagnostic Network Architecture

The architecture utilized for the diagnostics neural networks that follow was essentially a synthesis of the simple diagnostic model and the component isolating upper level network described above. In essence, the simple diagnostic model succeeded in providing both location and severity information on its own in that it provided a severity indication for a frequency or other parameter associated with a particular component based on a dB difference as an input. Its only drawback was that it was auto-associative, having the same number of inputs as outputs. The upper level network possessed hetero-associative characteristics in that the number of inputs differed from that of the outputs. The only other difference between the two preliminary networks was that one provided a non-linear mapping of a series of inputs with a comparatively wide variation of values into a series of outputs varying from 0.0 to 1.0, whereas the other received such a series of outputs. If the input to each PE in the input layer was normalized with respect to all other inputs to that PE so that the inputs were provided equal weight at the start of training, the need for the lower level network could be eliminated and both location and severity indicating tasks

could be combined in one network. The MinMax Table provides for this.

Neuralware Inc's MinMax routine is a simple algorithm which, prior to the presentation of a matrix of training vectors, scans the matrix by columns, picks out the maximum and minimum value, and normalizes all other intermediary values with respect to them. These normalized values are then retained in this normalized state or mapped linearly to another range of values at the discretion of the operator[Ref.8].

All of the prototype neural networks presented utilized the cumulative delta rule modification to the standard backpropagation algorithm described in Chapter II. They also utilized learning coefficients that decreased in steps as a function of the total number of training presentations. All processing elements in the hidden and output layer utilized the sigmoidal transfer function, while the input layer utilized a linear transfer function. No F' offset or momentum term was necessary.

The epoch size utilized in the cumulative delta rule varied from between 58 and 62 and the number of vectors in the training sets varied between 60 and 69. The slight deviation of the epoch size from the number of vectors in the training sets was intended to keep the sequence of training presentations between updates of the weights as varied as possible.

In the Neuralworks Professional II backpropagation routine, the order of test set presentation can be sequential or randomized immediately prior to training at the operator's discretion. In general it is desirable to present these vectors randomly. However, during the training process the training vectors are only randomized once. Thus the order of presentation does not change. If the number of training vectors is identical to the epoch size, the network updates the weights time after time on the same ordered presentation of vectors. If the epoch size is kept slightly less than the number of vectors in the training set, the network will update not having seen the entire training set. The following set of vectors presented to the network will pick up where the last epoch left off, considerably improving the variety of training vector sets presented to the network.

Schematics of the prototype diagnostics networks are provided in Figures 39, 40, and 41. The prototype diagnostic network each consisted of from 18 to 25 PE's in the input layer, 27 PE's in an hidden layer, and 7 PE's in the output layer. The outputs corresponded to the machinery component experiencing the fault and consisted of the high speed shaft (S1), the low speed shaft (S2), the high speed bearing inner race (BI), the high speed bearing outer race (BO), the bearing balls (BB), the 15 tooth drive pinion (G1), and the 50 tooth driven gear (G2).

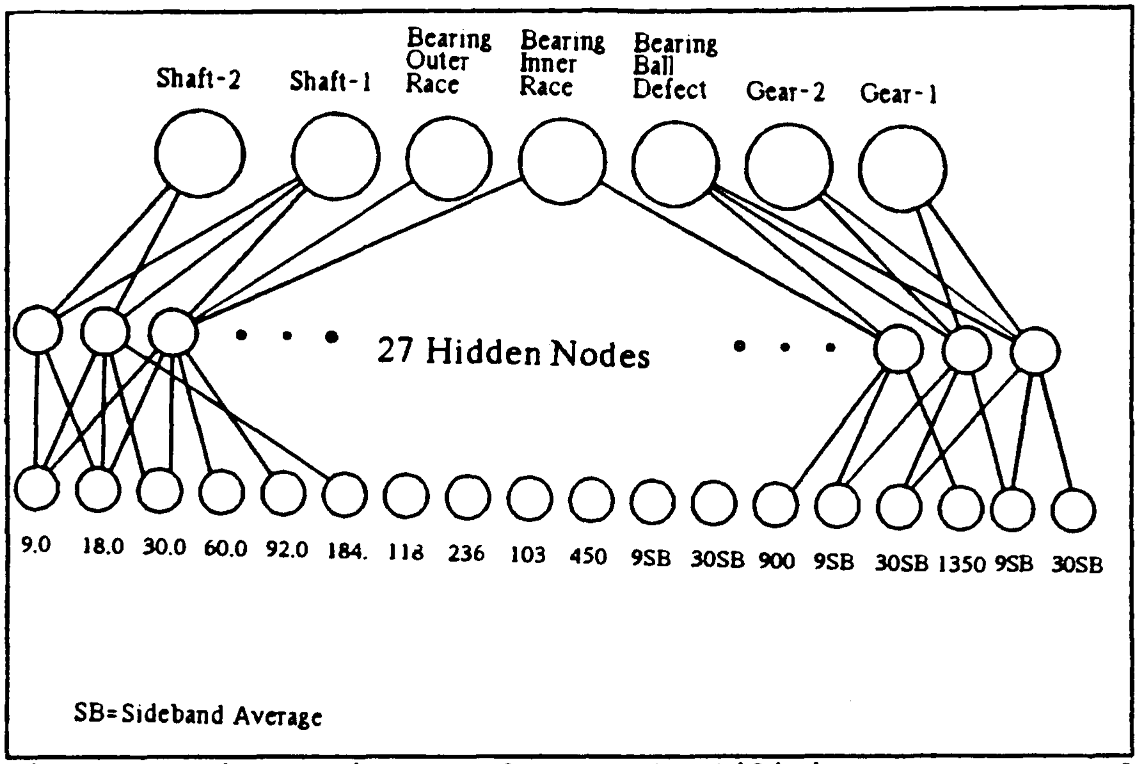

Figure 39 Diagnostic Neural Network Utilizing Frequency and Sideband Averaging Inputs

The inputs were limited to the dB levels for the frequencies and quefrencies monitored throughout the data extraction period. Three neural networks were developed. The first one employed purely frequency domain inputs and included the four frequencies corresponding to the shafts, five bearing frequencies, the three gear mesh frequencies, and the averages of the first three sidebands on each side of each of the gear mesh frequencies, totaling 18 inputs. A schematic of this network is provided in Figure 39. In the second network, the six sideband averaging inputs were replaced with three cepstral inputs associated with the gears and four cepstral inputs associated with the bearings, totaling 19

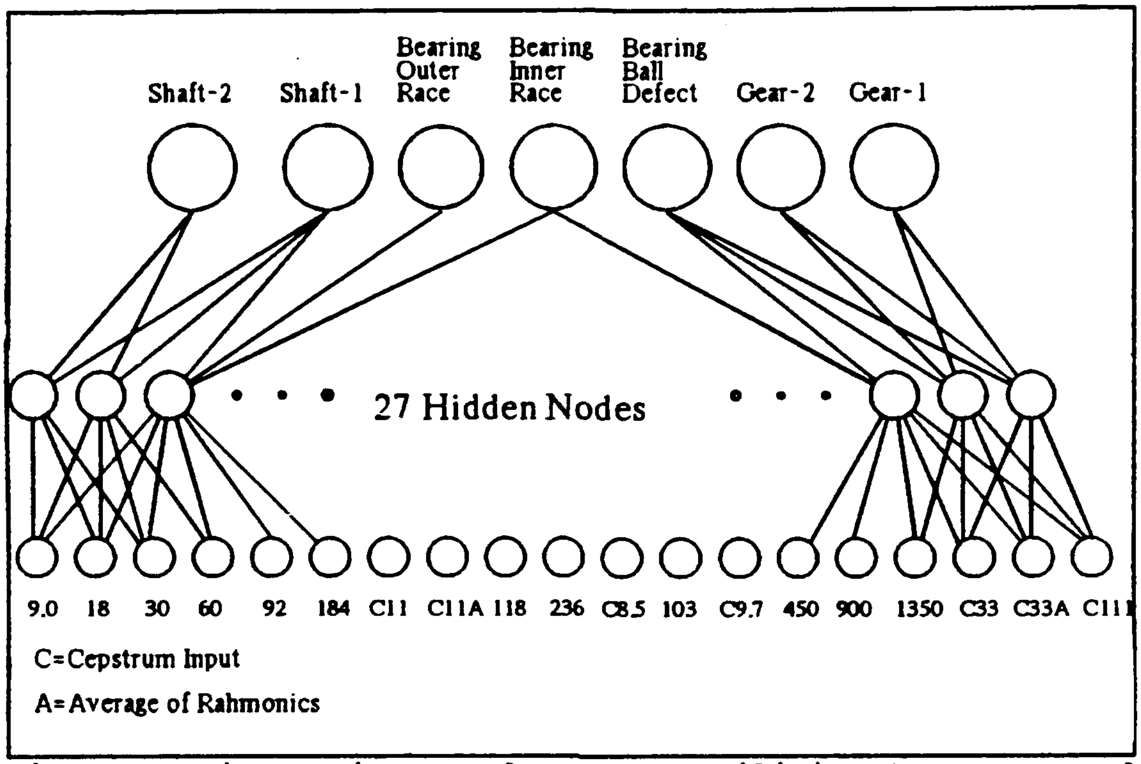

Figure 40 Diagnostic Neural Network Utilizing Frequency and Cepstral Inputs

inputs. Figure 40 illustrates this network. The third network utilized all monitored frequencies and quefrencies for a total of 25 inputs and is illustrated in Figure 41.

The initial number of hidden elements was determined by interpolating the results of the network sensitivity studies described in Chapter IV. Here it was determined that six hidden elements reached a 15% convergence level before any of the other networks studied and exhibited a high degree of stability as the error level declined. While Networks possessing fewer hidden elements also achieved convergence, they took longer to reach the 15% convergence level. Those with greater than six hidden elements became increasingly

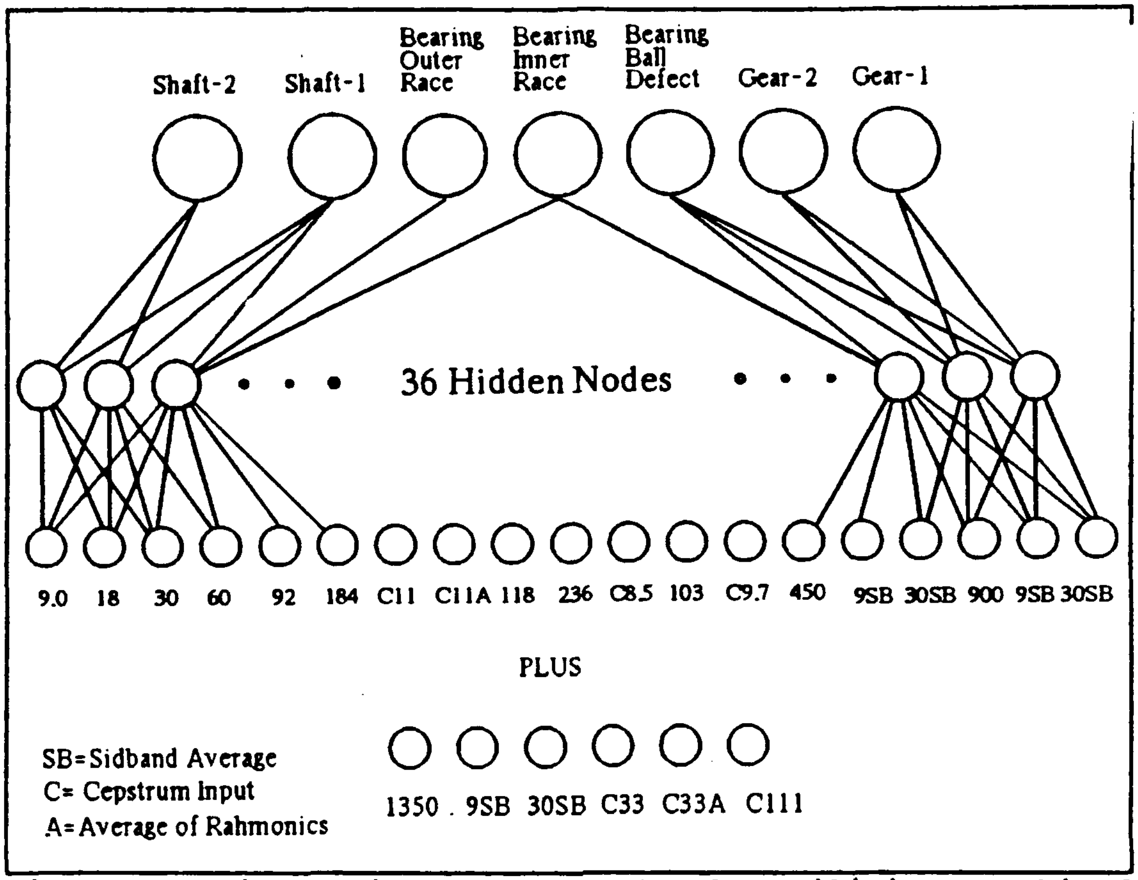

Figure 41 Diagnostic Neural Network Utilizing Combined Frequency, Cepstrum, and Sideband Averaging Inputs

unstable with respect to output error as the number of hidden elements increased. Since the number of hidden elements in the six hidden element network was 1.5 times the number of input elements and the input data was similar, the initial number of hidden elements in the prototype networks was determined to be 1.5 times the number of input elements. Thus the number of hidden elements for the sideband averaging and cepstrum networks was set at 27, while the number of hidden elements in the combined network was set initially at 38. Additionally, to reduce the computational burden of a large number of

connections with negligible excitation, a "prune network" feature was used. This feature permanently sets inactive connection weights to zero if after a given number of training iterations, the maximum activation energy fell below a set level. In the networks under consideration this parameter was checked every 10,000 iterations and the maximum activation threshold was set at 0.05[Ref.8]. This would appear to be a rather conservative figure as, during the training process for these networks, no connections were "pruned".

B. DESCRIPTION OF DATA SETS

The nature of the data sets utilized for training is critical to the success of a practical neural network based machinery diagnostic system. Especially important is the nature of any preprocessing done to the data prior to its input into the neural network. Clearly, a neural network's task in recognizing patterns can be made easier and thus, successful convergence of the error function can occur more quickly if the engineer's knowledge about the data base can be incorporated in the inputs before learning takes place. A possible danger also lies in incorporating too much a priori knowledge in that the neural network will be overly constrained, thereby losing the opportunity to identify relationships in the data that may not have been noted by the engineer.

For this research two different types of training sets were utilized for each of the prototype machinery diagnostics neural networks proposed. The first type of training set used consisted of 69 input vectors that were generated artificially, based on long established associations between certain frequencies and quefrencies and machinery faults. The second type of training set was extracted directly from empirical data obtained from the set of experiments described in Chapter V. Additionally, test sets containing data similar to that found in the parent training sets but nonetheless unique were built.

In this section, a detailed description of the data sets is provided. Details common to all data sets utilized are discussed first, followed by those aspects unique to each particular type of data set.

1. General Considerations

There were a number of considerations common to all data sets generated for use on the neural networks. A number of preprocessing steps were included to simplify the problem presented to the networks. Other preprocessing steps were accomplished because the networks simply appeared to have excessive difficulty solving the problem without the preprocessing.

In a manner patterned after the Navy surface ship machinery condition monitoring program, all measurements were

reduced to dB differences relative to an established baseline. Additionally, all data was passed through the MinMax normalization routine described above prior to being entered into the networks. While the raw dB values would have been normalized in the same manner as the dB difference values, their singularly negative values appeared to impose an excessive burden to the neural network without any significant return. Furthermore, expression of the values as differences from a baseline had the advantage of allowing the operator to recognize the relative strength of the signal at a glance and was in keeping with current practice. Thus the dB difference input form was retained.

Several attempts were made to train a network featuring training data with signed dB differences. As the sign of the dB difference had a major impact on the contribution of that particular input to the overall diagnosis, recognition of the sign of the input was highly desirable. Initial attempts involved lower level severity indicating networks with a signed input and an unsigned severity rating output. These were attempted with both sigmoid and hyperbolic tangent transfer functions. Follow on attempts exploited the positive and negative ranges of the hyperbolic tangent featuring both signed inputs and outputs. None of these variations provided a satisfactory convergence.

Initially the positive nature of the sigmoid transfer function was blamed for the difficulty. However, when it was

determined that the hyperbolic tangent transfer function was similarly unsuccessful, it was conjectured that the source of the problem lay in the way in which the backpropagation algorithm calculated error. Because it calculated the mean squared value of the error, the sign information was lost, thus leading to the difficulty.

Training sets that either truncated negative differences to zero or utilized absolute valued dB differences were considered. However, truncated dB differences were expected to significantly reduce the effectiveness of the gear mesh frequencies which often experienced reductions in dB level in cases of gear related faults. Absolute valued dB differences were expected to give unwarranted weight to lower frequency signals from the shafts and bearings which often declined in cases of gear faults. As a compromise, it was decided to enter negative dB differences into the training sets as zeros except for the gear mesh frequencies, where the absolute values were taken.

2. Artificially Generated Data Sets

Of the two types of data sets constructed, the artificially generated data set was by far the more difficult. Two training data sets were constructed, one for the network including sideband averaged inputs and one for the network including cepstral inputs. Each contained 69 input vectors.

These data sets presented data that was generated following established rules in machinery diagnostics. In these sets the system components were assumed to be essentially uncoupled, and the location of machinery faults was assumed to follow the following "rules".

- If the machine had elevated dB levels at the frequencies of 9 Hz or 18 Hz, the fault was assumed to be located at the low speed shaft (S2).
- If the machine had elevated dB levels at the frequencies of 30 or 60 Hz, the high speed shaft was the source of the fault (S1).
- If the machine had elevated dB levels at the frequencies of 92 or 184 Hz, or at the quefrency of 10.9 ms or the averaged 10.9 ms rahmonics, a fault existed at the outer race of the bearing (BO).
- If the machine had elevated dB levels at the frequencies of 118 or 236 Hz, or at the quefrency of 8.5 ms, a fault had occurred at the bearing inner race (BI).
- If the machine had elevated dB levels at 103 Hz or at a quefrency of 9.7 Hz, the fault was located at one of the bearing balls (BB).
- If the machine experienced elevated or depressed dB levels at the gear mesh frequencies of 450, 900, or 1350 Hz, a fault existed in one of the two gears or both.
- If the machine experienced elevated dB levels in any of the averaged 9 Hz sideband inputs, or at a quefrency of 111 ms, a fault existed in the 50 tooth gear (G2).
- If the machine had elevated dB levels in any of the averaged 30 Hz inputs, or at a quefrency of 33.3 ms, a fault existed on the 15 tooth gear (G1).
- If the magnitude of all associated inputs to a particular component were beneath their established low severity fault thresholds, no fault existed for that component.

Severity levels were established in a manner similar to that described in Chapter IV except that in this model, the severity level thresholds were based on the standard deviations measured in the baseline establishing experiments described in Chapter V. The "Low" severity fault level was based on the propagation of error formula for standard deviations using all baseline experiments."Moderate" and "High" severity fault levels were obtained by adding one or two of the highest standard deviations for that parameter in the three test sets, respectively to the "Low" severity threshold. This procedure is in keeping with most vibration monitoring manuals which indicate that a machinery fault can be expected to exist if the signal exceeds two standard deviations, which corresponds to a severity level between the low and moderate severities established in this research[Ref.27]. A listing of the severity thresholds used is provided in Table XI.

In preliminary experiments severity levels were established by devoting at least two training vectors to establish the high and low boundaries for all parameters. However, the networks trained in this manner had difficulty in discerning the boundary and, like the networks described in Chapter IV, performed poorly in the immediate area of the severity thresholds. By training in this manner, the network was unduly constrained and forced to accurately identify setpoints, a task where the essentially analog neural network

Table XI. Severity Thresholds for Artificially Generated Data Sets

Frequency (Hz)	9.0	18.0	30.0	60.0	92.0	103	118	184	236
First Threshold	4.5	4.5	4.1	2.5	4.1	2.0	3.6	6.6	5.9
Subsequent Thresholds	3.5	3.4	2.0	2.5	2.7	2.0	3.0	2.6	3.1
No Fault Median	2.3	2.3	2.1	1.3	2.1	1.0	1.8	3.3	3.0
Low Severity Median	6.3	6.2	5.1	3.8	5.5	3.0	4.8	7.9	7.5
Moderate Sev Median	9.8	9.6	7.1	6.3	8.2	5.0	7.1	10.5	10.6
High Severity Median	13.3	13.0	9.1	8.8	10.9	7.0	9.4	13.1	13.7

Frequency (Hz)	450	9SB	30SB	900	9SB	30SB	1350	9SB	30SB
First Threshold	4.1	3.2	3.0	2.0	2.0	2.3	3.2	2.9	2.0
Subsequent Thresholds	2.0	2.0	2.0	2.0	2.0	2.0	3.3	2.3	2.0
No Fault Median	2.1	1.6	1.5	1.0	1.0	1.2	1.6	1.5	1.0
Low Severity Median	5.1	4.2	4.0	3.0	3.0	3.3	4.9	4.1	3.0
Moderate Sev Median	7.1	6.2	6.0	5.0	5.0	5.3	8.2	6.4	5.0
High Severity Median	9.1	8.2	8.0	7.0	7.0	7.3	11.5	8.7	7.0

Cepstrum (ms)	8.5	9.7	10.9	10.9Av	33.3	33.3Av	111
First Threshold	0.5	1.0	0.5	0.4	0.9	0.7	1.0
Subsequent Thresholds	0.5	1.0	0.5	0.4	0.9	0.7	1.0
No Fault Median	0.3	0.5	0.3	0.2	0.5	0.4	0.5
Low Severity Median	0.8	1.5	0.8	0.6	1.4	1.1	1.5
Moderate Sev Median	1.3	2.5	1.3	1.0	2.3	1.8	2.5
High Severity Median	1.8	3.5	1.8	1.4	3.2	2.5	3.5

is categorically inefficient. It also forces a network of continuous transfer functions to provide a step output, another difficult task.

Better results were achieved in the prototype networks by concentrating training of the networks on the middle value of the desired severity region as opposed to the threshold.

Once the center of the severity region was established, the continuous nature of the transfer functions in the PE's would allow for interpolation of deviations from these median values. In essence, the constraints in the preliminary networks were relaxed and the network was allowed to establish the severity boundaries on its own, having the centers of the regions fixed instead. The difference in the means of defining the decision space is illustrated for a two dimensional case in Figure 42.

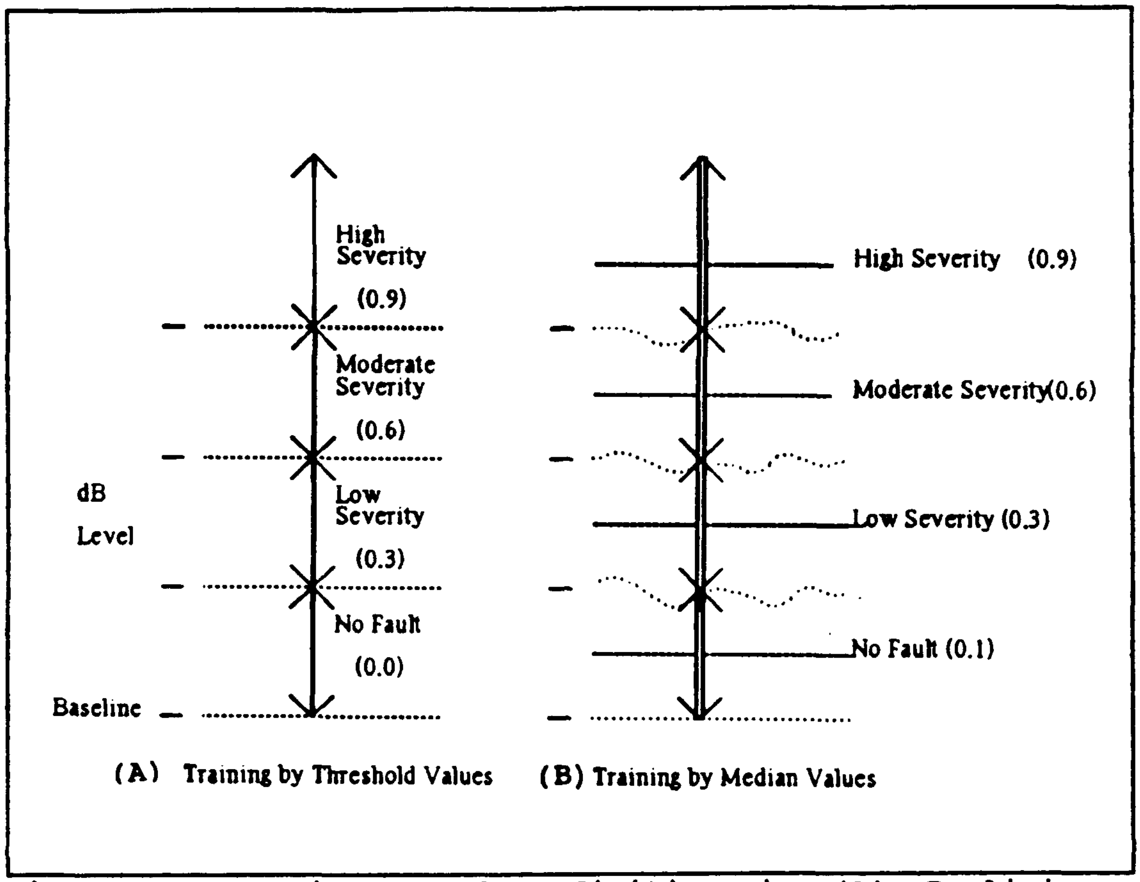

Figure 42 Severity Level Definition in (A) Preliminary Networks and (B) Prototype Networks

To further fix the centers of each severity region, the networks were only trained using the midpoint values at the desired severity region. Only in the training sets defining the no fault region were variations from these middle values permitted.

The desired outputs of the vectors in these data sets were established according to the procedures established in Chapter IV; that is, with outputs of 0.3, 0.6, and 0.9 corresponding to the three severity levels. Because of the fact that median values in each severity level were being used

in training, the desired output assigned for machinery components experiencing no faults was 0.1 vice the 0.0 output assigned in Chapter IV.

The test sets involved deviations about the mean severity level values, thereby providing for unique but similar vectors to those presented in the training set. In addition, a few new vectors, requiring a variation in the desired severity level output were included. The test sets contained a total of 63 vectors each.

3. Empirical Data Sets

The empirical data sets were comparatively easy to establish. All vectors were acquired by calculating the dB difference between the measured parameters and the established baselines. Half of the preprocessed vectors from each test set were used in the training sets while the other half were used in the test sets.

The severity criterion used in these sets was based on the assessment of the degree of physical damage discussed in Chapter V. If there was no fault associated with a particular machinery component, it was assigned a desired output severity level of zero. Clearly, there despite continuous pains to minimize it there was still some degree of mismatch between the severity criteria in the artificially generated and empirical data sets. This difficulty would manifest itself in

the results involving networks trained on artificially generated data but tested on empirical data.

C. PRESENTATION AND DISCUSSION OF RESULTS

A total of five prototype networks were trained and tested. The cepstrum and sideband averaging networks were each trained and tested on both artificially generated and empirical data, while the combined cepstrum and sideband averaging network was only trained and tested on empirical data. This section will describe and discuss the results of these tests. Additionally the results of a few tests stemming from networks trained on slightly erroneous data will be discussed. These erroneous test sets are included because they provide an insight to the robustness of the neural network as well as emphasizing the importance of verifying the correctness of the training set before training commences. Because the networks trained on erroneous data were subsequently trained on corrected data sets without starting afresh, their follow-on performance yields insight into the backpropagation neural network's capability to be updated as the data base changes over time. Before a discussion of the results is made, an explanation of how these results were derived is in order. A "correct diagnosis" was considered to have occurred if the network correctly identified the location of the fault, if there was one, or correctly identified no fault to exist if there was not. If the network correctly

identified the fault but also identified a lesser fault somewhere else whose severity level was sufficiently close to that of the principal fault so as to be possibly misconstrued to be the primary fault, a potential misdiagnosis was deemed to occur, which was treated as "50% correct". Additionally, in cases of multiple faults, the failure to identify any one faulty location while correctly identifying the principal fault and any other lesser faults was determined only to be a potential misdiagnosis. Blatant misdiagnoses were of course treated as such.

Severity error refers to the difference between desired output and actual output. Each vector was assigned to one of the four severity regions according to its highest severity error. When considering severity error, it must be remembered that the networks trained to artificially generated data were trained to median severity values. Thus a severity error of between 15 to 25 percent is not necessarily an unexpected or bad thing. However, errors greater than 25 percent should be regarded with some degree of suspicion. Most of the cases of blatant misdiagnosis stem from severity errors greater than 25 percent but in some cases, especially where low severity levels were involved, a potential misdiagnosis or even a blatant misdiagnosis could and did occur with errors as low as 10 percent.

In the sections to follow, tables are used to summarize the test results. Included in the tables is a distribution of

severity errors among the seven outputs. When viewing this severity distribution data for the empirical test and training sets , it must be borne in mind that the bulk of the empirical data obtained involves gear related faults. Consequently several of the other components are only stimulated a few times. Thus while it may appear that the Shaft 1 output generated less error than Gears 1 and 2, it actually performed less well when not reporting a no fault condition.

1. Network with Sideband Averaging Inputs

a. Artificially Trained Network Response

The sideband averaging network was first trained using artificially generated data to an RMS error level of 0.065 after 355,374 presentations of the training data set. This network was subsequently tested on the data set it was trained on, a separate artificially generated test set, and on a test set containing empirical data. A summary of the results of these tests is provided in Table XII.

Of the five prototype networks trained, the sideband averaging network "learned" its training set best, succeeding in correctly identifying 100 percent of the faults and determining the severity level to within 20 percent in almost 90 percent of the cases. The network performance when presented the artificially generated test set resulted in only a 4.0 percent degradation in correct diagnoses. Severity level error only degraded by 6.0 percent. However, the network

Table XII. Artificially Trained Sideband Averaging Network Test Response

		Artificial Training Set		Artificial Test Set		Empirical Test Set	
Correct Diagnosis		100%		96.4%		77.5%	
< 20% Error		89.8%		83.8%		28.3%	
< 15% Error		78.3%		70.5%		25.0%	
< 10% Error		56.5%		44.1%		16.7%	
Location of Severity Errors, Artificial Training Set							
Error	S2	S1	BO	BI	BB	G2	G1
>20%	0	1	1	2	0	1	1
15-20	1	2	2	2	0	2	3
10-15	4	2	3	1	2	4	1
< 10%	64	64	63	64	67	62	64
Severity Error Location For Artificial Test Set							
> 20%	0	2	2	6	0	2	3
15-20	2	1	2	1	1	1	3
10-15	2	3	2	4	3	6	7
< 10%	64	62	62	57	64	59	55
Severity Error Location For Empirical Test Set							
> 20%	7	10	7	11	8	18	23
15-20	1	5	1	2	4	1	6
10-15	2	2	2	6	4	2	6
< 10%	50	43	50	41	44	39	25

performance on the empirical data test set was disappointing. Only 77.5 percent of the test vectors were successfully

diagnosed, while only 28 percent of the test cases had all severity level errors less than 20 percent.

The network trained on artificially generated data responded well when presented with artificially generated test data. However, faults where a single input provided the only indication of the fault were consistently underestimated. This is not overly surprising when considering the manner by which the network output is attained. Furthei, most prudent machinery diagnostics "experts" look at a fault identified by a single high parameter with a jaundiced view, tending to verify the calibration of the particular instrument before taking corrective action.

The rather disappointing network response to empirical data can be partially explained by noting that the rules under which the network was trained did not account for the coupling between the various machinery components. Even if it had been included, it would not have been expected that the coupled component would register a higher severity level than the component experiencing the fault. This was precisely what occurred in several of the shaft related faults. Although the networks were unable to identify the shaft as the source of the fault, they did faithfully register faults on the components whose associated inputs received high dB levels, which was what the network was trained to do. Another situation glaringly evident from the network response to empirical test data was the fact that increased dB level is

not the only determinant in the severity of physical damage to the equipment.

b. Empirically Trained Network Response

The sideband averaging diagnostic neural network trained on a data set of empirical data reached a level of 0.075 RMS error after 187,218 iterations. A summary of the test results for the empirically trained sideband network is presented in Table XIII. When tested on the same data it performed at a level only 4.8 percent below that of the artificially trained network tested on the training set. When tested on new data the network suffered a significant degradation but a general diagnosis success rate and severity error rate 11.4 percent better and 82.6 percent better, respectively than that of the artificially trained network tested on empirical data. While severity level accuracy in the declined by 40.2 percent between the tests on the training data and the previously unseen data, fault location capability remained fairly high, degrading by only 8.5 percent.

Notable areas of weakness were in detecting the high speed shaft faults and identifying weak faults on Gear 1 when coupled with severe shaft faults. Another area of weakness lay in identifying borderline low severity Gear 1 faults in the data extracted from Gear Test 1-4 described in Chapter V.

Table XIII. Summary of Empirically Trained Sideband Averaging Network Test Results

	Training Set			Test Set			
Successful Diagnosis	95.2%			86.7%			
Severity Error < 20%	91.9%			51.7%			
Severity Error < 15%	74.2%			41.7%			
Severity Error < 10%	61.3%			30.0%			
Empirical Training Set Severity Error Location							
Error	S2	S1	BO	BI	BB	G2	G1
> 20%	0	3	0	0	0	1	1
15-20	0	0	0	0	0	1	10
10-15	0	0	1	1	1	2	7
< 10%	62	59	61	61	61	58	44
Empirical Test Set Severity Error Location							
> 20%	0	3	2	2	2	10	16
15-20	0	0	0	0	0	3	6
10-15	0	0	0	0	0	1	10
< 10%	60	57	58	58	58	46	28

2. Networks With Cepstral Inputs

a. Network Trained on Artificially Generated Data

The Cepstral network was tested after reaching an RMS error of 0.068 after 663,000 iterations, of which 250,000 occurred after correcting a minor error in the training set.This network performed in a manner similar to that of

their sideband network counterparts. This network was very successful in determining the location of the machinery faults on the artificially generated training set and test set, successfully diagnosing the location of these faults in all but one case. Furthermore, severity errors were the smallest found in any of the networks tested. However, the test response to presentation of the empirical test set was considerably less successful than in the case of the sideband networks, primarily due to a paucity of cepstral information provided in the cases where both Gear 1 and Gear 2 were damaged (Gear Test 1-10) and in several cases involving damage to Shaft 2, where strong 33.3 ms cepstral signals mislead the network into identifying Gear 2 as the source of the fault. Additionally, because of elevated 30 or 60 Hz signals in the more severe gear faults, the network tended to downplay the importance of these signals. Table XIV provides a summary of these results.

b. Empirically Trained Network

The empirically trained version of the cepstral network was tested upon achieving an RMS error of 0.095 after 532710 iterations. Like its artificially trained counterpart, it performed poorly on gear faults involving Gear 2 where 111 ms cepstrum input did not register. The other place where this network performed poorly was on faults involving Shaft 1, where, because of elevated 30 or 60 Hz signals in the more

Table XIV. Test Results for Cepstrum Network Trained on Artificially Generated Data

		Artificial Training Set		Artificial Test Set		Empirical Test Set	
Correct Diagnosis		99.3%		100%		61.7%	
< 20% Error		85.5%		84.1%		26.7%	
< 15% Error		79.7%		68.1%		21.7%	
< 10% Error		60.8%		56.5%		15.0%	
Error	S2	S1	BO	BI	BB	G1	G2
Artificial Training Set Severity Error Location							
>20%	1	2	4	2	2	1	2
15-20	0	1	2	0	0	1	0
10-15	4	2	2	2	2	4	2
<10%	64	64	61	65	65	63	65
Artificial Test Set Severity Error Location							
>20%	3	2	3	4	4	1	1
15-20	0	4	2	3	2	2	1
10-15	2	2	1	1	1	1	4
<10%	64	61	63	61	62	65	63
Empirical Test Set Severity Error Location							
>20%	7	6	8	16	2	17	17
15-20	2	3	1	3	4	4	2
10-15	1	0	2	1	3	7	3
<10%	50	51	49	40	51	32	38

severe gear faults, the network tended to downplay the importance of these signals. However, overall, the empirically trained network performed quite well compared to

the artificially trained network tested on empirical data, outperforming the artificially trained network by 32.3 percent in fault location identification and by 68.5 percent in severity error. Its performance was slightly less impressive than that of the empirically trained sideband averaging network, successfully diagnosing 6.0 percent fewer test vectors and possessing a 6.0 percent higher severity error, but its performance was comparable. A summary of the results of these tests are presented in Table XV.

c. Erroneous Training Sets

During the training of the prototype networks, cepstral networks were inadvertently trained on data sets which contained one or two clerical errors among the 60 or more vectors involved which degraded these sets' utility with respect to establishing an effective machinery diagnostics system. They were subsequently retrained with corrected data sets. However, the limited manner by which these errors degraded the test response lends insight into the robustness of the neural networks and their tolerance to noisy data. Because of this, their test response is also reported.

The cepstrum network trained on noisy artificial data was tested after reaching an RMS error of 0.085 after 409,371 iterations. Surprisingly enough, the cepstrum networks trained on this slightly faulty data performed almost as well as the networks trained on correct data. The results of these

Table XV. Test Results for Empirically Trained Network with Cepstral Inputs

		Training Set			Test Set		
Correct Diagnosis		93.5%			81.7%		
<20% Error		88.7%			45.0%		
<15% Error		71.0%			31.6%		
<10% Error		53.2%			26.7%		
Error	S2	S1	BO	BI	BB	G1	G2
Severity Error Location for Empirical Training Set							
>20%	0	3	0	0	0	3	2
15-20	0	0	0	0	0	11	3
10-15	1	1	0	0	0	8	5
<10%	61	58	62	62	62	40	52
Severity Error Location for Empirical Test Set							
>20%	0	3	0	0	0	24	12
15-20	1	0	0	0	0	7	5
10-15	3	2	0	0	0	2	3
<10%	56	55	60	60	60	22	40

tests, compared to their counterparts trained on correct data are presented in Table XVI.

Because the errors involved in the training set were relatively minor, it was decided to simply continue training using the corrected training set rather than reinitializing the network and starting training afresh. Although the errors to the training set occurred in the input vectors, the desired output was altered in the corrected

Table XVI. Comparison of Cepstral Network Trained on Slightly Faulty Data and Correct Data

	Correct Diag.	< 20% Error	< 15% Error	< 10% Error
Faulty Training Set	97.1%	82.6%	72.5%	59.4%
Correct Training Set	99.3%	85.5%	79.7%	60.8%
Faulty Test Set	95.6%	75.4%	68.1%	56.5%
Correct Test Set	100%	84.1%	68.1%	56.5%

training set to speed up learning, as this would produce strong error signals directly, rather than allowing the change to be filtered through the entire network. Indeed, observation of the RMS error immediately after continuing training revealed a substantial increase in RMS error which eventually subsided, confirming the effectiveness of the approach. Fortunately the errors occurred in the artificially trained network. Had they occurred in the empirically trained network, this method would not have been appropriate.

In the previous case, the error involved an error in the inputs which altered the severity level required at a desired output from a 0.1 to a 0.3 and another one from a 0.3 to a 0.6. The next case involves a considerably more severe clerical error, where the location of a high severity fault was shifted from Gear 1 to Gear 2 in one sample vector. Here reinitialization of the network was considered prudent due to the magnitude of the error. The effect of the error was to

suppress the severity levels experienced by the component where faults were actually occurring but the whose desired output indicated a no fault condition and to amplify the low dB signals associated with the component which in reality was experiencing no fault at all. In spite of this error which confused the network somewhat, the network still was capable of performing quite well, exceeding the performance of networks trained on artificially generated data on empirical test sets. A summary of these test results are provided in Table XVII. Interestingly, the network trained on erroneous data actually performed about 6.0 percent better than the empirical test set than did the network trained on correct data.

These two examples, inadvertently happened upon, serve to demonstrate the robustness of the neural network diagnostic system. It is doubtful that a rule based expert system would have been able to perform as well with conflicting data. The first example also demonstrates the ability for the network to update its data base without having to start training from scratch.

3. Combined Sideband and Cepstrum Diagnostics Network

Because of the paucity of cepstral information in the empirical data on several of the faults involving both Gears 1 and 2, as well as difficulties in identifying faults involving Shaft 1, a machinery diagnostics neural network

Table XVII. Comparison of Networks Trained on Erroneous and Corrected Empirical Data

	Correct Diagnosis	< 20% Severity Error	< 15% Severity Error	< 10% Severity Error
Erroneous Training Set	91.9%	75.8%	61.3%	41.9%
Correct Training Set	95.2%	91.9%	74.2%	61.3%
Erroneous Test Set	83.3%	61.7%	48.3%	36.7%
Correct Test Set	86.7%	51.7%	41.7%	30.7%

combining both cepstral and sideband averaging inputs was built, trained, and tested. Only empirical data was used as there was no difficulty in training and testing the previous two networks on artificially generated data. This network was tested after 444,981 iterations of the training set and achieving an RMS error of 0.09. Test results are presented in Table XVIII.

Compared with the sideband averaging network trained on empirical data, the combined network performed equally well when determining location of the faults and had improved by approximately 13 percent in severity error. When tested on the empirical test set it performed 1.7 percent better in fault location and 9.4 percent better in severity the accuracy of its severity indication.

Table XIII. Test Results for Combined Network Trained on Empirical Data Sets

		Empirical Training Set			Empirical Test Set		
Correct Diagnosis		95.2%			88.3%		
< 20% Error		95.2%			60.0%		
< 15% Error		90.3%			51.7%		
< 10% Error		85.5%			40.0%		
Error	S2	S1	BO	BI	BB	G1	G2
Severity Error Distribution: Empirical Training Set							
>20%	0	3	0	0	0	0	0
15-20	0	0	0	0	0	3	0
10-15	0	0	0	0	0	3	0
<10%	62	59	62	62	62	56	62
Severity Error Distribution: Empirical Test Set							
>20%	0	3	0	0	0	12	9
15-20	0	0	1	1	1	4	2
10-15	1	0	0	0	0	8	2
<10%	59	57	59	59	59	36	47

Comparison to the cepstral network performance on empirical data yields even more impressive results. When responding to the training set, the combined network outperformed the cepstral network by 1.7 percent in fault location and 19.7 percent in severity accuracy. Combined network response against the empirical test set was also impressive. It outperformed the cepstral network by 6.7 percent in fault identification and by 16.1 percent in

severity accuracy. However, even with all data obtained from the experiments performed, the fault to Shaft 1 could not be identified, indicating that the shaft frequency signals were too small for recognition compared to the considerably larger gear vibrations also in progress.

4. Results of Extended Learning

All network training and testing conducted up to this point was conducted on either an IBM 286 or 386 personal computer. On the 386 computer,also equipped with a math coprocessor, neural networks with the number of PE's of the order utilized in this research commonly required 12 hours to conduct 200,000 training iterations. Very late into this research, a Unix SUN Spark station became available. The cepstrum network and its associated empirical data training set were loaded and run on this station overnight for 4.5 million training iterations using the standard backpropagation algorithm. At this length of training the RMS error was reduced to 0.01 and the response to the training set resulted in 100 percent successful fault location and 100 percent of the severity determinations remaining at less than 10.0 percent error.

D. EVALUATION OF EMPIRICAL INPUTS

In this section an analysis of the relative effectiveness of the inputs selected for the neural networks will be made. As a whole, judging from the overall effectiveness of the

various networks, it would appear that the inputs encompassed the decision space for the networks fairly well with a few notable exceptions. None of the three networks adequately identified the actual fault experienced by the high speed shaft. This could be either due to shortcomings in the data set or in the inputs themselves. Proper determination of this would require expanding the data set to incorporate additional examples of shaft and bearing faults. Additionally, based on the cepstral network response to empirical data, it would appear that for the machine studied, cepstral inputs alone were insufficient to identify faults involving both gears, since sideband and combined networks were able to correctly diagnose these faults.

A good source of insight into the relative effectiveness of the various inputs may lie in observing what inputs were important to the empirically trained networks following their long periods of training. While theoretically, the information by which the neural network separates the decision space can be found in the hidden PE's, the source of most feature extraction. However, thus far no knowledge has been amassed as to how this knowledge might be extracted[Ref.19].

A more primitive and less comprehensive alternate means to obtain a feel for the relative importance of the various inputs may come from sequentially stimulating input neurons (processing elements) and observing the resulting output, much like a doctor checking nervous reflexes. This was attempted by

constructing a test data set which was constructed of vectors that provided a maximum input to one input node while providing zeros to all of the others. This methodology was applied to all three of the empirically trained networks. They reveal some startling results.

Table XIX. Combined Network Response to Sequential Input Neuron Stimulation

Freq (Hz)	9.0	18.0	30.0	60.0	92.0	184	C10.9	C10.9A	118	236	C8.5	103	C9.7
S2	54.9	80.4	0.7	0.0	5.0	8.4	17.9	9.4	0.2	2.3	1.8	1.9	39.6
S1	0.1	0.0	0.0	0.0	0.2	0.0	0.2	0.4	0.0	0.0	0.0	0.0	0.3
BO	0.2	0.0	0.0	0.0	4.8	0.5	2.8	0.4	0.1	0.4	0.8	0.4	0.3
BI	0.2	0.0	0.0	0.0	5.3	0.6	3.2	13.7	0.1	0.5	0.8	0.4	0.4
BB	0.2	0.0	0.0	0.0	5.1	0.6	3.1	13.8	0.1	0.5	0.9	0.4	0.4
G1	10.3	19.3	94.8	3.0	70.6	29.6	68.1	34.0	31.1	1.0	0.2	0.9	45.0
G2	34.1	0.3	4.4	96.9	9.2	60.2	4.6	28.7	68.3	95.1	95.7	79.5	14.1
Total Norm.*	1.6	5.0	9.7	10.1	4.8	0.7	1.5	2.8	6.6	2.3	3.9	2.7	0.4

* Total output of the network due to a specific neural stimulation normalized by 10 0 over the sum of the outputs due to the neural stimulations

Freq (Hz)	450	9SB	30SB	900	9SB	30SB	1350	9SB	30SB	C33	C33A	C111	
S2	0.2	36.7	1.1	65.9	0.3	26.0	1.1	70.0	3.5	8.8	14.4	84.3	
S1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.0	
BO	0.3	0.1	1.3	0.0	0.1	0.0	0.1	0.0	0.0	6.1	1.7	0.3	
BI	0.3	0.2	1.3	0.0	0.1	0.0	0.2	0.0	0.0	6.0	1.6	0.4	
BB	0.3	0.2	1.4	0.0	0.1	0.4	0.2	0.0	0.0	7.5	1.9	0.5	
G1	80.5	3.9	75.6	2.1	80.7	74.0	1.2	27.5	96.1	17.4	83.4	0.0	
G2	18.5	58.8	19.2	31.8	18.5	0.0	97.2	1.5	0.3	53.9	11.3	14.5	
Total Norm.*	13.1	3.0	3.7	1.9	3.4	8.8	4.4	0.5	1.7	1.0	0.8	5.4	

The results of the neural stimulation test on the trained combined network are summarized in Table XIX. The four inputs that appear to have been used least by the network were the 184 Hz signal, the 9 Hz sidebands about 1350 Hz, and the 9.7 ms and 33.3 ms average cepstral inputs. By far the greatest bulk of the output activation occurred in those

output neurons that received the greatest overall stimulation throughout the training process; that is, the two gears. Certain inputs were used by the neural network to provide strong output signals to components that they were not directly associated with. The more notable associations include those linking the 60 Hz signal to Gear 2, the bearing frequency and cepstral domain signals to either or both of the gears, particularly for 92 and 184 Hz and their related cepstral signals; and the 111 ms cepstrum to Shaft 1. The network frequently linked gear related cepstral inputs to the corresponding shafts, which is understandable. Randall[Ref.34] indicates that in broad band cepstra the low frequencies associated with the shafts often affect the quefrencies associated with the gears and thus suggests that a band pass filter be utilized to cut the low frequencies out. Finally there is the very noticeable fact that shaft 1 received no significant activation from any of the inputs.

The other two networks performed in a similar manner to that observed in the combined network. One notable exception is that the Shaft 1 output in the cepstrum network is considerably more strongly represented than in either the sideband averaging network or the combined network. Presumably the elimination of all output energy from Shaft 1 is derived from the sideband averaging network. A coarse summary of these test results is provided in Table XX.

Table XX. Cepstrum and Sideband Network Responses to Sequential Neuron Stimulation

Sideband Averaging Network Response To Neuron Stimulation Test										
Frequency (Hz)	9.0	18.0	30.0	60.0	92.0	184	118	236	103	
Output Signal Strength (Norm to100)	1.3	7.8	13.7	4.4	3.8	1.1	5.9	1.3	3.5	
Principle Location of Signal	S2,S1	S1	G1	G2	B	B	G2	B	B	
Secondary Location(s) of signal	B	(S2)	B	B	G1,G2	S1	G1	G2	G2	
Frequency (Hz)	450	9SB	30SB	900	9SB	30SB	1350	9SB	30SB	
Output Signal Strength (Norm to 100)	10.4	4.9	9.0	7.5	1.5	7.2	5.6	0.7	2.9	
Primary Location of Signal	G1	G2	G1	S2	G1	G1,S2	G2	S2	G1	
Secondary Location of Signal	(G2,B)	(B,S1)	B,S1	G2	S1,B	(S1)	(S1)	S1,B	(S1)	
Cepstrum Network Response To Neuron Stimulation Test										
Frequency (Hz)	9.0	18.0	30.0	60.0	92.0	184	C10.9	C10.9A	118	236
Output Signal Strength (Norm to100)	4.2	5.6	10.4	4.2	4.1	1.7	1.7	4.2	9.5	2.2
Principle Location of Signal	S2	S2	G1	G2	G1	G1	G1	B	G2	G2
Secondary Location(s)	(S1,G1)	(S1,G1)	(S1*)	G1 (S1)	B,S1	S1	S1	S1	G1	S1
Frequency (Hz)	C8.5	103	C9.7	450	900	1350	C33.3	C33.3A	C111	( )= less than 15% of total output signal
Output Signal Strength (Norm to 100)	6.8	8.7	4.4	10.9	3.7	3.0	4.1	1.4	10.0	
Primary Location of Signal	G2	G1	G1	G1	S2	S1	G1	S1	S2	
Secondary Location of Signal	(S1)	(S2,S1)	(S1)	(G2)	(G1,G2 S1)	G2(S2)	S1	G1	G2	(*) less than 5%

The results of this section are not definitive. The effects of multiple combined inputs, transfer functions,and wide ranging connection weights have not been considered. The purpose of this section is merely to gain a crude insight as to the relative effectiveness of the various inputs. The empirically trained networks still provide a diagnostics

capability on real data that is consistently superior to that provided by the artificially trained networks. What these results do bring out is the probability that a wider data base consisting of a larger proportion of shaft and bearing faults may yield better results and a confirmation that 92 and 184 Hz may have been confused from time to time with the much more dominant shaft rotation harmonics of 90 and 180 Hz. In spite of this possibility, the networks performed remarkably well in detecting the location and severity of the limited number of bearing faults imposed during the experimental data extraction phase of this research.

VII SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

A. SUMMARY OF RESULTS

In preliminary experiments described in Chapter IV, a rudimentary neural network architecture for machinery diagnostics utilizing the historically successful backpropagation algorithm was established. These simple four input/four output networks were capable of determining the location and severity of faults in between 85 and 90 percent of the test vectors presented after training on artificially generated data over less than 80,000 iterations. During these experiments an optimal number of hidden nodes for that particular network and type of training data was determined to be between four and eight, with the six hidden node network reaching an initial level of convergence in the least number of vector presentations.

Following this, a data base was established for an uncomplicated gear train system with multiple machinery components by observing the vibration signatures at discrete points in the frequency spectrum and cepstrum associated with the machinery components of interest. After establishing a baseline using undamaged components, machinery faults were imposed and the system response was observed.

The results from these experiments are discussed in detail in Chapter V but the principal results were as follows. In general, the physical system responded as would be expected according to well established rules of machinery diagnostics. However, the system experienced a larger degree of coupling among machinery components and increases in physical damage were found not always to result in increases in vibration level.

While empirical data was still being obtained, the prototype neural networks were being developed. These networks were similar in architecture to the ones developed in the preliminary experiments but were larger, hetero-associative, and utilized the cumulative delta rule with sigmoid transfer functions vice the normalized cumulative delta rule and hyperbolic tangent transfer functions utilized in the preliminary experiments. Additionally the prototype neural networks utilized a linear mapping algorithm to normalize the various inputs. Severity levels were established based on the standard deviations observed at each input parameter during baseline tests for use in artificially generated training sets and based on engineering judgement for the empirical training sets.

Three networks were developed; one using sideband averaging inputs to assist in gear fault diagnostics, one using ceptral inputs to aid in diagnostics of bearing and gear faults, and one combining both sideband averaging and cepstral

inputs with frequency domain inputs. Two of these prototype networks were first trained and tested on artificially generated data based on the established rules of machinery diagnostics. These networks successfully diagnosed the fault location for almost 100 percent of the sample vectors present in the artificially generated training and test sets and succeeded in keeping error in severity level below 20 percent in 84 percent of the sample vectors presented. These tests included multiple faults. When presented with empirical data, correct diagnosis dropped to an average of 68 percent of the test vectors and severity errors under 20 percent dropped to a mere 27 percent. This was due to the strong coupling between machinery components and the nonlinearities involved in the correlation between severity level and vibration magnitude. Cepstral networks performed slightly less well than sideband averaging networks, presumably due to the reduced range of dB values experienced in the cepstrum.

Then all three prototype networks were trained and tested on empirical data. The networks were able to correctly diagnose the location of the fault and kept severity error below 20 percent in an average of 94.6 percent and 91.9 percent of the vector presentations, respectively. When presented with the empirical test sets they averaged 85.6 percent for successful location diagnosis and 52.2 percent for severity error less than 20 percent. While this is a significant drop from the training set it is a substantial

improvement over the empirical test results obtained from the artificially trained networks. Of the three networks, the combined network displayed the best performance while the cepstrum network performance was least impressive.

Principal causes of the errors were a paucity of cepstral information in the multiple gear fault cases, the indirect relationship between dB level and physical damage, and the consistent failure to identify faults associated with the high speed shaft. The reasons for the third cause involve misleading rises in the frequencies and quefrencies associated with the high speed pinion, but more importantly, the tendency for the shaft rotative frequencies to become elevated during gear faults which tended to drive down the sensitivity of all networks to faults involving the high speed shaft.

Late into the research, a SUN station became available for limited use. After 4.5 million training presentations from the empirical training set using the standard bacpropagation algorithm, the network was able to correctly identify all faults and correctly diagnose the severity level for all vectors presented to within ten percent.

B. CONCLUSIONS

Based on the results cited above as well as in the body of this paper, the following conclusions may be drawn.

All neural networks trained on actual and artificially generated data demonstrated a capacity for simultaneous

multiple fault detection, an area where conventional expert systems have commonly fallen short.

Based on the results from the preliminary experiments and the response of the artificially trained and tested networks, it is clear that neural networks utilizing the architecture noted in this paper are capable of being successfully trained and tested on artificially generated data reflecting the established rules of machinery diagnostics.

Disappointing results experienced with the artificially trained networks tested on empirical data indicate that the rules utilized for training did not adequately account for the strong inter-component coupling associated especially with small, light weight mechanical systems.

From the empirical data as well as the results from the testing of the artificially trained networks on empirical data, it is also clear that dB level and severity of physical damage, while related, are not directly proportional.

Neural networks utilizing the architecture described in this paper and trained on empirical data are capable of reaching exceptional levels of convergence given sufficient training as evidenced by the cepstrum training on the SUN station. At less extreme lengths of training, these same neural networks can achieve an acceptable level of convergence.

Inasmuch as the network trained for an extensive period was able to reach an exceptional level of convergence, it is

clear that, for the data set acquired, the inputs utilized were sufficient to describe the decision space. However, the failure of the empirically trained networks to successfully identify faults to the high speed shaft at less extreme lengths of training indicates that an investigation into providing additional inputs or expanding the data base to incorporate additional shaft and bearing fault information would prove prudent.

Cepstrum networks inadvertently trained on artificially generated and empirical data tainted with minor errors suffered only a slight degradation of performance. This demonstrates that neural networks of the architecture described possess an inherent robustness and tolerance to noisy data not generally found in conventional expert systems.

Finally, empirically trained networks consistently outperformed artificially trained networks when tested on empirical data. This indicates that the neural network was able to discern both the non-linear relationship between dB level and severity of physical damage and the coupling relationships between machinery components. While by no means comprehensive, the neuron stimulation tests clearly implied that some of the relationships between frequencies and their related components had changed. The artificially trained network, in reality a rule based expert system by reason of the method by which it was "taught", was incapable of learning these relationships because they were not in the rule base.

This demonstrates an inherent advantage of the data based learning of the neural network over the rule based learning of the conventional expert system.

C. RECOMMENDATIONS FOR FURTHER STUDY

The research presented in this paper is by no means complete. There remains a large number of areas for additional study. Some of the many areas recommended for further expansion include the following.

The data base utilized for this research is by no means complete and warrants further expansion, particularly in the number of shaft and bearing faults imposed. Additionally, the data extracted in this research was generally obtained and processed manually and was therefore painfully time consuming. Automation of the data extraction, preprocessing and neural network interface processes would reduce the opportunity for error while increasing the number of faults that could be imposed dramatically. Furthermore, the small size of the machinery components enhanced the degree of coupling between components in the system and reduced the loading on the bearings to virtually nil. Because of this the gear vibrations predominated throughout the spectrum and tended to mask out the bearing vibrations. Increasing the size of the machinery components could go a long way in alleviating this problem.

The accuracy of the artificially generated data base may have been improved by employing a computational modal analysis

routine to predict the response of uncomplicated machinery to various faults. However, as the purpose of this research is to obtain a diagnostic system for complex machines well beyond the capabilities of current modal analysis techniques, this approach may be self defeating. Another approach might be patterned after the research conducted by Sejnowski and Rosenfeld[Ref.39] in speech generation where a neural network was trained using an existing rule based expert system[Ref.19]. In a similar manner, artificially trained diagnostic neural networks might be trained by an off-the-shelf rule based expert system might yield improved results.

There is still substantial work available in optimizing the network architecture. The two level network originally planned for implementation in this research had to be abandoned prematurely due to time constraints and a belatedly discovered correctable error after an alternative architecture capitalizing on the MinMax Table to replace the lower level networks was found to work satisfactorily. Inasmuch as the upper and lower level networks worked well independently, this architecture may have proven optimal.

A substantial amount of effort was spent on attempting to find a means by which to effectively train on signed inputs and desired outputs. In an effort to circumvent this problem, the data had to undergo additional preprocessing based on statistical observations. While this may have been a practical solution, information potentially useful to the network had to

be discarded. Research into this problem may also reap significant benefits.

This research primarily concentrated on the use of backpropagation as the learning algorithm of choice due to its historical success. Hcwever, although backpropagation has its place in machinery diagnostics, it is data intensive. Unfortunately, the data base available for most large expensive machines is limited at best. Furthermore, it is economically unfeasible to conduct destructive testing on the large, expensive pieces of machinery that would stand to benefit most from a machinery diagnostic system. Research into the use neural networks utilizing unsupervised learning algorithms such as the Adaptive Resonance Theory series under development by Grossberg may prove to be a more practical alternative.

This research has demonstrated that neural networks have a place in machinery condition monitoring and diagnostics. However the limited nature of these results indicate that neural networks will not solve all machinery condition monitoring and diagnostics problems by themselves. They certainly will not completely replace conventional rule based expert systems. Ultimately it is anticipated that a symbiotic combination of these two technologies will provide the optimal solution to the machinery condition monitoring and diagnostics problem.

APPENDIX A.

Sample Training and Test Sets Used in Preliminary Experiments.

Table A1. Sample Test Set Input and Output

INPUT				OUTPUT			
X1	X2	X3	X4	Y1	Y2	Y3	Y4
0.0	0.0	1.5	0.5	-0.0450	-0.0255	0.0063	-0.0033
0.0	0.2	0.7	0.5	-0.0457	-0.0217	-0.0215	-0.0018
1.8	1.2	0.7	0.5	0.0396	0.0098	-0.0237	-0.0172
0.2	1.4	1.7	0.3	-0.0411	0.0126	0.0181	-0.0072
0.8	1.7	0.9	0.7	-0.0247	0.0409	-0.0172	0.0001
2.5	1.2	1.4	1.0	0.2198	0.0097	-0.0002	-0.0139
0.7	2.8	1.2	0.9	-0.0262	0.2251	-0.0079	0.0117
0.9	3.2	1.8	2.1	-0.0249	0.3524	0.0281	0.1129
1.0	2.0	3.1	0.5	-0.0125	0.0498	0.3491	-0.0100
0.8	1.5	1.8	2.1	-0.0334	0.0199	0.0295	0.1124
4.7	2.2	1.1	0.8	0.6809	0.1157	-0.0196	-0.0479
0.5	2.8	0.7	4.2	-0.0718	0.2537	-0.0263	0.7299
0.8	2.8	5.3	0.3	-0.0131	0.2341	0.7457	-0.0110
6.2	2.2	1.1	0.8	0.8201	0.1177	-0.0198	-0.0521
1.1	6.9	1.8	2.1	-0.0111	0.8609	0.0269	0.1019
2.5	2.1	6.2	0.7	0.2073	0.0392	0.7926	-0.0297

Table A2. Input and Desired Output of Training Set

INPUT				DESIRED OUTPUT			
X1	X2	X3	X4	Y1	Y2	Y3	Y4
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0
1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
2.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0
2.0	2.0	1.0	1.0	0.0	0.0	0.0	0.0
2.0	2.0	2.0	1.0	0.0	0.0	0.0	0.0
2.0	2.0	2.0	2.0	0.0	0.0	0.0	0.0
2.0	1.0	2.0	1.0	0.0	0.0	0.0	0.0
1.0	2.0	1.0	1.0	0.0	0.0	0.0	0.0
2.5	1.0	1.0	1.0	0.3	0.0	0.0	0.0
2.5	2.0	1.0	0.0	0.3	0.0	0.0	0.0
1.0	2.5	1.0	0.0	0.0	0.3	0.0	0.0
1.0	2.0	2.5	1.0	0.0	0.0	0.3	0.0
1.0	1.0	1.0	2.5	0.0	0.0	0.0	0.3
3.0	1.0	1.0	1.0	0.3	0.0	0.0	0.0
1.0	3.0	2.0	1.0	0.0	0.3	0.0	0.0
1.0	1.0	3.0	1.0	0.0	0.0	0.3	0.0
2.0	1.0	2.0	3.0	0.0	0.0	0.0	0.3
3.0	2.0	3.0	2.0	0.3	0.0	0.3	0.0
3.5	2.0	3.5	3.0	0.3	0.0	0.3	0.3
1.0	2.0	3.5	1.0	0.0	0.0	0.3	0.0
4.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0

Table A2A. Sample Training Set Inputs and Desired Outputs(cont.)

INPUT				DESIRED OUTPUT			
X1	X2	X3	X4	Y1	Y2	Y3	Y4
2.5	1.0	4.0	1.0	0.3	0.0	0.6	0.0
2.0	5.0	1.0	0.0	0.0	0.6	0.0	0.0
2.5	2.0	2.0	5.5	0.3	0.0	0.0	0.6
1.0	2.0	5.0	1.0	0.0	0.0	0.6	0.0
5.0	1.0	2.0	1.0	0.6	0.0	0.0	0.0
5.0	3.0	2.0	1.0	0.6	0.3	0.0	0.0
3.0	4.0	2.0	1.0	0.3	0.6	0.0	0.0
2.0	2.0	3.0	4.0	0.0	0.0	0.3	0.6
5.5	3.0	4.0	3.0	0.6	0.3	0.6	0.3
1.0	3.0	4.0	2.0	0.0	0.3	0.6	0.0
6.0	1.0	2.0	1.0	0.9	0.0	0.0	0.0
3.0	6.5	1.0	1.0	0.3	0.9	0.0	0.0
2.0	1.0	6.0	1.0	0.0	0.0	0.9	0.0
1.0	2.5	1.0	6.0	0.0	0.3	0.0	0.9
2.0	7.0	1.0	0.0	0.0	0.9	0.0	0.0
1.0	1.0	1.0	6.0	0.0	0.0	0.0	0.9
7.0	3.0	2.0	2.0	0.9	0.3	0.0	0.0
2.0	7.0	2.0	3.0	0.0	0.9	0.0	0.3
2.0	3.0	2.0	6.0	0.0	0.3	0.0	0.9
6.0	4.0	3.0	1.0	0.9	0.6	0.3	0.0
1.0	3.0	4.0	6.0	0.0	0.3	0.6	0.9
3.0	2.0	7.0	2.0	0.3	0.0	0.9	0.0

LIST OF REFERENCES

1. Watanabe, K.,Matsuura, I., Masahiro, A., Kobota, M., and Himmelbau,D.M., "Incipient Fault Diagnosis of Chemical Processes via Artificial Neural Networks", American Institute of Chemical Engineers Journal, Vol. 35 No. 1, November 1989, pp 1803-1812.

2. Venkatasubramanian, V., and Chan, K., "A Neural Network Methodology for Process Fault Diagnosis", American Institute of Chemical Engineers Journal, Vol. 35 No. 12, December 1989, pp 1993-2002.

3. Porenta, G., Dorfner, G., Schedlemayer, J., and Sochor, H., "Parallel Distributed Processing as a Decision Support Approach in The Analysis of Thallium-201 Scintigrams", Proceedings of the IEEE Conference on Computers in Cardiology, 1988, pp 259-262.

4. Iwata, A., Nagasaka, Y., Suzumura, N., "A Digital Holter Monitoring System with Dual 3 Layers Neural Networks", Proceedings of the IEEE International Joint Conference on Neural Networks, Vol. II, 1989, pp 69-74.

5. Marko, K.A., James, J., Dosdall, J., Murphy, J., "Automotive Control System Diagnostics using Neural Nets for Rapid Pattern Classification of Large Data Sets", Proceedings of the IEEE International Joint Conference on Neural Networks, Vol. II, 1989, pp 13-16.

6. McDuff, R., Simpson, P., Gunning, D., "An Investigation of Neural Networks for F-16 Fault Diagnosis: I. System Description", Proceedings from IEEE AutoTestCon, 1989, pp 351-357.

7. Dietz, W.E., Kiech, E.L., and Ali, M., "Jet and Rocket Engine Fault Diagnosis in Real Time", Journal of Neural Network Computing, Vol. 1, Summer 1989, pp 5-18.

8. Klimasauskas, C., Guiver, J., Pelton, G., Neural Computing Neuralworks Professional II Operator's Manual, Neuralware Inc., Pittsburgh, Pennsylvania, 1989.

9. McCulloch, W.S., and Pitts, W., "A Logical Calculus of the Ideas Immanent in Nervous Activity", Bulletin of Mathematical Biophysics, 1943, Vol. 5, pp 115-133.

10. Hebb, D.O., The Organization of Behavior, Wiley, New York, 1949.

11. Rosenblatt, F., "The Perceptron: a Probablistic Model for Information Storage and Organization in the Brain", Psychological Review, Vol. 65, 1958, pp 386-408.

12. Widrow, B., and Hoff, M., "Adaptive Switching Circuits", 1960 IRE WESTCON Convention Record. IRE, New York, pp 96-104.

13. Anderson, J., Rosenfeld, E., Ed., Neurocomputing: Foundations of Research, MIT Press, Cambridge, Massachusetts, 1988.

14. Minsky, M., Pappert, S., Perceptrons, MIT Press, Cambridge, Massachusetts, 1969.

15. Werbos, P., Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD Thesis, Harvard University Committee on Applied Mathematics. November 1974.

16. Grossberg, S., "How Does a Brain Build a Cognitive Code", Psychological Review, Vol. 87, 1980, pp 1-51.

17. Hopfield, J.J., "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", Proceedings of the National Academy of Sciences, Vol. 79,1982, pp 2554-2558.

18. Rummelhart, D.E., McClelland, J.L., Ed. Parallel Distributed Processing: Explorations in the Microstructures of Cognition Vol.1: Foundations, MIT Press, Cambridge, Massachusetts, 1986.

19. Hecht-Nielsen, R., Neurocomputing, Addison-Wesley, New York, 1990.

20. Gorman, R.P., Sejnowski, J.J., "Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets", Neural Networks, Vol.1, 1988, pp 75-89.

21. Allred, L.G., "Neural Networks in Automotive Testing of Diode Protection Circuits", Proceedings of 1989 IEEE AutoTestCon, 1989, pp 181-183.

22. Fahlmann, S.E., "An Empirical Study of Learning Speed in Back Propagating Networks", CMU Technical Report, CMU-CS-88-162, June 1988.

23. Dubose, P., and Klimausauskas, C., "An Introduction to Neural Networks with Examples and Applications", Neural Network Short Course Notes, Neuralware Inc. , Pittsburgh, Pennsylvania, 1989.

24. Carpenter, G. and Grossberg, S., "A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine", Computer Vision, Graphics, and Image Understanding, Vol.37, 1987, pp 54-115.

25. Mark, W.D., "Gear Noise Origins", AGARD Conference Proceedings, No.369, pp 30-1-30-14.

26. Matthew, J., and Alfredson, R.J., "The Monitoring of Gearbox Vibration Operating Under Steady Conditions", Proceedings of the 11th Biennial Conference on Mechanical Vibration and Noise, Mechanical Signature Analysis, Boston, Massachusetts, September 1987, pp 47-54.

27. Hewlett Packard, Dynamic Signal Analyzer Applications: Effective Machinery Maintenance Using Vibration Analysis, Application Note 243-1, Hewlett Packard, 1983.

28. Randall, R.B., "A New Method of Modeling Gear Faults", ASME Journal of Mechanical Design, Vol.104, 1982, pp 259-267.

29. Favaloro, S.C., "A Preliminary Evaluation of Some Gear Diagnostics Using Vibration Analysis"(Australian) Department of Defence Science and Technology Organisation of Aeronautical Research Laboratories, Aero Propulsion Technical Memorandum 427, Melbourne, Australia, 1985.

30. Sandy, J., "Monitoring and Diagnostics for Rolling Element Bearings", Journal of Sound and Vibration, June 1988, pp 16-20.

31. Collacott, R.A., Vibration Monitoring and Diagnostics:Techniques for Cost Effective Plant Maintenance, Wiley, New York, 1979.

32. Matthew, J., "Monitoring the Vibrations for Rolling Machine Elements -An Overview", ASME Proceedings of the 1987 Biennial Conference on Mechanical Vibration and Noise, Boston, Massachusetts, September 1987, pp 15-22.

33. Dyer, D. and Stewart, R.M., "Detection of Rolling Element Bearing Damage by Statistical Vibration Analysis", ASME Journal of Mechanical Design, Vol.100, April 1978, pp 229-235.

34. Randall, R.B., "Advances in the Application of Cepstrum Analysis to Gearbox Diagnosis", Second International Conference on Vibrations in Rotating Machines, Cambridge, Institute of Mechanical Engineering, London, September 1980, pp 159-174.

35. VanDyke, D.J., and Watts, W.A., "Automated Rolling Contact Bearing Fault Detection Using Cepstrum Analysis" Proceedings of the Second International Machinery Monitoring and Diagnostics Conference, Society of Experimental Mechanics, Union College, Schenectady, New York, 1990, pp 554-559.

36. Strunk, W.D., "Considerations for the Establishment of a Machinery Monitoring and Analysis Program for Surface Ships of the U.S. Navy", Proceedings of the 6th International Modal Analysis Conference, Vol.II, Society of Experimental Mechanics, Union College, Schenectady, New York, 1988,pp 914-920.

37. Robinson, J.D., "Statistical Approach to Fault Detection of Gears", Master's Thesis, Naval Postgraduate School, Monterey, California, December 1989.

38. Matthew, J., and Alfredson, R.J., "The Condition Monitoring of Rolling Element Bearings Using Vibration Analysis", Journal of Vibration, Acoustics, Stress, and Reliability, Vol.106, July 1984, pp 447-453.

39. Sejnowski, T.J., and Rosenberg, C.R., "NETtalk: A Parallel Network that Learns to Read Aloud", Johns Hopkins University EE & CS Technical Report, JHU/EECS-86/01, January 1986.